arXiv:astro-ph/9607088v1 18 Jul 1996

CALCULATION OF COSMIC BACKGROUND RADIATION ANISOTROPIES AND IMPLICATIONS

EMORY F. BUNN
Astronomy Department
University of California, Berkeley

Abstract. We review the physical processes that are thought to produce anisotropy in the cosmic microwave background, focusing primarily (but not exclusively) on the effects of acoustic waves in the early Universe. We attempt throughout to supply an intuitive, physical picture of the key ideas and to elucidate the ways in which the predicted anisotropy depends on cosmological parameters such as Ω_0 and h. The second half of these lectures is devoted to a discussion of microwave background data analysis techniques, with an emphasis on the analysis of the COBE DMR data. In particular, the Karhunen-Loève method of data compression is described in detail.

1. Introduction

Since the discovery four years ago of cosmic microwave background (CMB) fluctuations (Smoot *et al.* 1992), the data from anisotropy experiments have improved in both quality and quantity at a very rapid pace. CMB data already provide stringent constraints on cosmological models, and with a plethora of balloon-borne and ground-based experiments underway and two planned satellite missions, we can expect further dramatic improvement over the next decade. In fact, there is a very real possibility that we will accurately measure many of the most important cosmological parameters via the CMB anisotropy spectrum (Jungman *et al.* 1996, Kosowsky *et al.* 1996).

In order to realize this promise, we must take great care in developing tools for comparing observational data with theoretical predictions. Even with existing data, this process is far from trivial, and with the much larger data sets of the near future the task will become trickier. There are at least two independent problems to be faced: we must be able to make accurate

predictions of the anisotropy spectrum for any particular theory, and we must develop adequate statistical techniques to facilitate the comparison of these predictions with observations.[1]

These lectures are concerned with these two subjects. We will first review the primary physical mechanisms that are thought to be responsible for generating CMB anisotropies. The emphasis in this half of the lectures will be on building an intuitive picture of the relevant physical effects. We will therefore give ourselves free rein to make physically motivated approximations, rather than trying to treat the rather involved subject of anisotropy formation with complete precision. This section of the lectures will draw heavily on the work of Wayne Hu and Naoshi Sugiyama (Hu & Sugiyama 1994, 1995a, 1995b, 1996; Hu 1995), as well as on a review article by Hu, Sugiyama, & Silk (1996) and two previous summer-school proceedings on the subject (Hu 1996, Tegmark 1996c).

The second half of these lectures is devoted to issues of statistics and data analysis. We will study various ways in which theoretical predictions of CMB anisotropy may be compared with data sets. Our primary focus will be on methods for analyzing the COBE DMR data, since this is the largest and most powerful CMB data set in existence; however, many of the issues that arise in analyzing the COBE data are directly relevant to analyses of other experiments, both present and future. For example, we will pay special attention to the issue of *data compression*; this subject was fairly important in analyzing the COBE data, and its importance will only increase as CMB data sets get larger and larger. In particular, the planned MAP and COBRAS/SAMBA missions will both return data sets several orders of magnitude larger than COBE, and their analysis will therefore require extensive data compression.

These lectures are organized as follows. Section 2 provides an overview of the key physical processes that produce CMB anisotropy. Section 3 discusses the primary anisotropy, including the Sachs-Wolfe effect (Sachs & Wolfe 1967) and anisotropies produced by acoustic oscillations of the photon-baryon fluid (Peebles & Yu 1970; Doroshkevich, Zel'dovich, & Sunyaev 1978; Bond & Efstathiou 1984), as well as the diffusive damping of fluctuations (Silk 1968). In Section 4 we discuss anisotropies produced after last scattering, such as the integrated Sachs-Wolfe effect (Sachs & Wolfe 1967, Rees & Sciama 1968), the effect of gravitational lensing (Blandford & Narayan 1992, Seljak 1996b), and reionization (Sunyaev 1977, Silk 1982). Section 5 attempts to synthesize the main ideas of the previous sections and concludes the first half of these lectures.

[1]Not to mention the far more difficult task of actually gathering the data!

The second half, which concerns issues of statistics and data analysis, begins with Section 6, in which we establish some basic results and notation having to do with Gaussian random processes on the sphere. Section 7 presents a series of idealized thought experiments designed to introduce some of the key issues of CMB data analysis. This section also contains a digression on Bayesian and frequentist statistical techniques. In Section 8, we apply what we have learned to an analysis of the four-year COBE DMR data, and Section 9 contains some brief concluding remarks.

2. An Overview of Anisotropy Formation

CMB anisotropies encode large amounts of information about the Universe. Physical processes around the redshift of last scattering (typically $z \simeq 1100$) produce the *primary anisotropy*, which can be significantly altered by *secondary* processes between the last-scattering surface and the present. In addition, the angular scale subtended by a particular source of anisotropy depends on the spatial geometry as well as the distance to the last-scattering surface.

With the exception of some effects at very low redshift, and ignoring topological defect models, calculations of CMB anisotropy are done in linear perturbation theory. All of the relevant quantities are small perturbations about a homogeneous Friedmann-Robertson-Walker solution. Nonetheless, making accurate numerical predictions of the CMB anisotropy in a particular theory is a daunting numerical task. In a typical cold dark matter (CDM) model, the variables one must keep track of include

- $\delta_{\rm B} \equiv \delta\rho_{\rm B}/\rho_{\rm B}$, the baryon density perturbation.
- $\delta_{\rm CDM} \equiv \delta\rho_{\rm CDM}/\rho_{\rm CDM}$, the perturbation in the CDM density.
- $\mathbf{v}_{\rm B}$, the baryon peculiar velocity field.
- $\mathbf{v}_{\rm CDM}$, the CDM peculiar velocity field.
- Ψ, essentially the Newtonian gravitational potential.
- Φ, the perturbation to the spatial curvature.[2]
- f_γ, the photon phase-space distribution function.
- f_ν, the neutrino phase-space distribution function.

All of these quantities depend on position $\mathbf{x}$ and time t, and f_γ and f_ν are also momentum-dependent. Their evolution is governed by a nasty set of coupled partial differential equations. For the nonrelativistic species, we must keep track of the usual equations of perturbation theory, namely the

[2]We will work throughout in Newtonian gauge. For our purposes Ψ and Φ are the only important perturbations to the metric. Ψ is related to the perturbation to the time-time component g_{00} of the metric, and Φ has to do with the perturbation to the spatial part g_{ij}. For more information on gauges, see the contribution of J.-L. Sanz to this volume, and also Hu (1995, 1996) and references therein.

continuity equation, the Euler equation, and the Poisson equation. For the CDM, these equations look like

$$\dot{\delta}_{\rm CDM} + \nabla \cdot \mathbf{v}_{\rm CDM} = 0, \tag{1}$$

$$\dot{\mathbf{v}}_{\rm CDM} + 2\frac{\dot{a}}{a}\mathbf{v}_{\rm CDM} = -\frac{1}{a^2}\nabla\Psi, \tag{2}$$

$$\nabla^2\Psi = 4\pi G\bar{\rho}\delta. \tag{3}$$

Here a is the scale factor, $\bar{\rho}$ is the average density, and a dot denotes a time derivative. All spatial derivatives are taken with respect to comoving coordinates. In the last equation, δ represents the total density perturbation, although we will generally consider models that are gravitationally dominated by CDM, so that we can replace δ with $\delta_{\rm CDM}$. There are also continuity and Euler equations for the baryons, the latter containing a pressure term.

The relativistic species (photons and neutrinos) are not characterized by a simple velocity field, but by a distribution function whose evolution is governed by the Boltzmann equation,

$$\frac{Df}{Dt} \equiv \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x^i}\frac{dx^i}{dt} + \frac{\partial f}{\partial p}\frac{dp}{dt} + \frac{\partial f}{\partial \gamma^i}\frac{d\gamma^i}{dt} = C[f]. \tag{4}$$

Here p is the magnitude of the momentum, γ^i is a direction cosine of the momentum, and C is a collision term having to do with scattering. This equation applies to both f_γ and f_ν, although at the epochs we are interested in the neutrino collision term is zero.

In order to make accurate predictions of the CMB anisotropy in a particular model, it is necessary to solve this system of equations numerically. If we work in Fourier space, we find that different fluctuation modes are uncoupled and the solution is therefore greatly simplified. We write

$$\delta(\mathbf{x}, t) = \sum_{\mathbf{k}} \delta_{\mathbf{k}}(t) \exp(i\mathbf{k}\cdot\mathbf{x}), \tag{5}$$

and similarly for the other quantities. [For the distribution functions, it is convenient to make a second expansion in Legendre polynomials $P_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{p}})$.] The fact that different $\mathbf{k}$-modes decouple makes the problem computationally tractable. Furthermore, as we shall see, the fact that we can work with one mode at a time makes it easier to get a conceptual understanding of anisotropy formation.

In recent years excellent codes have been developed for integrating these equations. [See Hu *et al.* (1995) and Bond (1996) for fairly recent discussions of the state of the art, and Seljak & Zaldarriaga (1996) for an important subsequent development.] We will not discuss the details of such precise

calculations here; rather, we will follow a less precise but more intuitive picture of the formation of anisotropies, based on a series of physically motivated approximations. This approach makes it easier to see what the important physical processes are and also gives us an understanding of how various features in the anisotropy spectrum depend on key cosmological parameters.

We will begin by discussing the sources of primary anisotropy: the Sachs-Wolfe effect (Sachs & Wolfe 1967), which describes gravitational red- and blueshifts due to potential differences on the surface of last scattering; the Doppler effect due to bulk motions of the last-scattering surface (Sunyaev & Zel'dovich 1970); and intrinsic temperature variations from point to point (Silk 1967). We will then discuss some sources of secondary anisotropy, the most important of which is the integrated Sachs-Wolfe (ISW) effect, which describes energy changes in photons as they pass through time-varying potentials. [This effect was also treated by Sachs & Wolfe (1967), as well as by Rees & Sciama (1968) at nearly the same time.] Other secondary sources of anisotropy include scattering by reionized matter and gravitational lensing.

At first, we will consider the evolution of only one Fourier mode at a time; however, we will eventually need to synthesize all of the different Fourier modes together to see what the total CMB anisotropy on the sky looks like. To do that, we will need to know the *power spectrum* of the density perturbation. This is simply the mean-square amplitude of the various Fourier modes:

$$P(k) = \langle |\delta_{\mathbf{k}}|^2 \rangle. \tag{6}$$

(As long as space is isotropic, P depends only on the magnitude of $\mathbf{k}$.) The angle brackets here denote an ensemble average, although it is frequently acceptable to assume δ is ergodic, in which case the angle brackets can equally well be regarded as a spatial average.[3] We often assume that the initial power spectrum is a power law in k: $P(k) \propto k^n$. As we will see below, the analogous quantity for describing the observed CMB anisotropy is the *angular power spectrum*:

$$C_l = \langle |a_{lm}|^2 \rangle. \tag{7}$$

Here a_{lm} is a coefficient of an expansion of a spherical harmonic expansion of the temperature anisotropy (spherical harmonic expansions being the natural analogue of Fourier expansions for data sets that live on the sphere). A mode with spherical harmonic index l probes an angular scale on the sky of $\theta \sim l^{-1}$. In any particular cosmological model, the angular power spectrum C_l is related linearly to the matter power spectrum $P(k)$. The

[3]Beware: When we describe $\Delta T/T$ as a random field on the sphere, we may *not* assume ergodicity: $\Delta T/T$ is never ergodic.

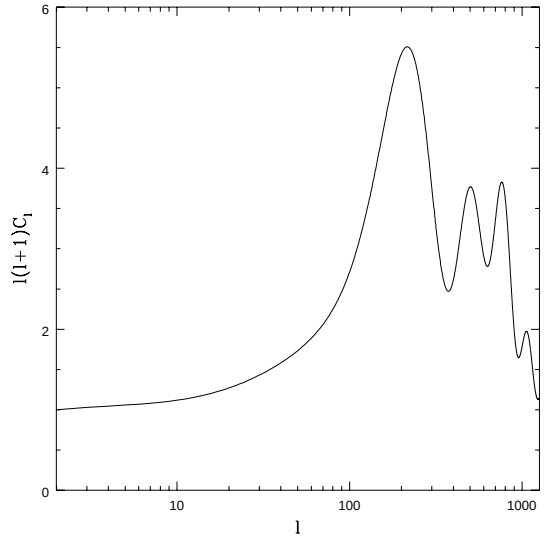

Figure 1. The angular power spectrum $l(l+1)C_l$ for a standard cold dark matter model. The parameters of this model are as follows: $n = 1$, $h = 0.5$, $\Omega_0 = 1$, $\Omega_{\rm B}h^2 = 0.013$. This power spectrum was computed by N. Sugiyama.

angular power spectrum for a CDM model is shown in Figure 1.[4] The primary goal of Section 3 will be to explain the multiple peaks in this spectrum.

3. Primary Anisotropies

3.1. THE GRAVITATIONAL POTENTIAL

We will begin by assuming that, after the end of the radiation epoch, most of the mass in the Universe is in the form of cold dark matter:

$$\Omega_{\rm B} \ll \Omega_{\rm CDM}. \tag{8}$$

Then the gravitational potential is completely determined by the CDM, and three equations (1−3) can be solved for Ψ and δ without worrying about what the other species are doing. Then, once we know the gravitational potential Ψ, we can solve for the evolution of the photons and baryons.

[4]The prefactor $l(l+1)$ in Figure 1 (and all of the other power spectrum plots we will see) is traditional. In a flat cosmological model with an $n = 1$ power spectrum, the Sachs-Wolfe contribution to the power spectrum is proportional to $1/l(l+1)$. The Sachs-Wolfe effect dominates on large scales, explaining the flatness of Figure 1 at low l. The quantity $l(l+1)C_l$ is also approximately proportional to the total power per logarithmic interval in l. (To make this proportionality exact, one would use $l(l+\frac{1}{2})C_l$ instead.)

Equations (1−3) can be combined into a single second-order equation for δ,

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} - 4\pi G\bar{\rho}\delta = 0. \tag{9}$$

At early times, when the Universe is radiation dominated, the last term in this equation is negligible, and the two linearly independent solutions are $\delta = \text{const.}$ and $\delta \propto \ln t$. There is therefore little growth during the radiation era.

If the Universe is matter dominated (meaning that both radiation and curvature are negligible in the Friedmann equation), then we have $a \propto t^{2/3}$, and the solutions are $\delta \propto t^{2/3} \propto a$ and $\delta \propto t^{-1}$. At late times, of course, the growing mode is the one that matters. If we plug the matter-dominated growing-mode solution into the Poisson equation (3), we find that Ψ *is independent of time*. This is a key fact, to which we will return repeatedly.

3.2. THE PHOTON-BARYON FLUID

Now that we know what the gravitational potential is doing, we are ready to study the evolution of the photons and baryons. We do this by making another approximation: we assume *tight coupling* between photons and baryons. Specifically, we assume that the mean free time τ between photon collisions is small compared to the other important time scales:

$$\tau \ll H^{-1}, (ck)^{-1}, (c_s k)^{-1}. \tag{10}$$

Here H^{-1} is the expansion time scale, $(ck)^{-1}$ is the light-travel time across a Fourier mode, and $(c_s k)^{-1}$ is the sound-travel time across a mode (c_s being the sound speed). This is an excellent approximation right up until around the time of last scattering.

In the tight-coupling approximation, frequent scattering isotropizes the photon distribution function f_γ: at any particular point, f_γ is isotropic in the rest frame of the baryons at that point. In fact, f_γ is completely characterized by the temperature distribution. Furthermore, the photon and baryon densities are coupled adiabatically: $n_\gamma \propto n_{\rm B} \propto T^3$. The behavior of the photon-baryon fluid is therefore characterized by a single variable: if we know, say, $\delta_{\rm B}(\mathbf{x}, t)$, we can determine $\mathbf{v}_{\rm B}$, T, and f_γ. We will find it convenient to take as our variable the fractional temperature fluctuation, which is simply one third of the baryon density fluctuation:

$$\Theta(\mathbf{x}, t) \equiv \frac{\Delta T}{T}(\mathbf{x}, t) = \frac{1}{3}\delta(\mathbf{x}, t). \tag{11}$$

With these approximations, the dynamics of the photon-baryon fluid is described by the single equation

$$\frac{d}{d\eta}\left[(1+R)\dot{\Theta}\right] + \frac{k^2}{3}\Theta = F(\eta). \tag{12}$$

This equation comes from the Euler and continuity equations for the fluid. We are working in units in which $c = 1$. For a derivation of this equation, see Hu (1995). In this equation, η is the conformal time,

$$\eta = \int^t \frac{dt}{a(t)}, \tag{13}$$

and $R \equiv 3\rho_{\rm B}/4\rho_\gamma$ is essentially the baryon-to-photon energy ratio. The overdot denotes a derivative with respect to conformal time. This equation is in Fourier space, so $\Theta = \Theta_{\mathbf{k}}$ represents a single Fourier mode with wavenumber $\mathbf{k}$.[5] The right-hand side $F(\eta)$ is a gravitational driving term,

$$F(\eta) = -\frac{k^2}{3}(1+R)\Psi - \frac{d}{d\eta}\left[(1+R)\dot{\Phi}\right]. \tag{14}$$

The rest of this section will be devoted almost entirely to a discussion of the solution of equation (12). We begin by making some useful observations. First,

$$R = \left(\frac{450}{1+z}\right)\left(\frac{\Omega_{\rm B}h^2}{0.015}\right), \tag{15}$$

where h is the Hubble parameter in units of $100\,{\rm km\,s^{-1}\,Mpc^{-1}}$, z is the redshift, and $\Omega_{\rm B}$ is the baryonic contribution to the density parameter. So for standard recombination at $z \simeq 1000$ and baryon densities around the nucleosynthesis range, $R \simeq \frac{1}{2}$ at the time of last scattering.

With the approximations that we're making, there are no anisotropic stresses, so the two gravitational potentials are simply related to each other:

$$\Phi = -\Psi. \tag{16}$$

Furthermore, we have seen that during the matter-dominated epoch, if linear theory is valid, Ψ is independent of time. The gravitational driving term therefore simplifies to

$$F(\eta) = -\frac{k^2}{3}(1+R)\Psi. \tag{17}$$

[5]It has become standard practice in cosmology to denote functions and their Fourier transforms by the same symbol, relying on context to tell the difference. [For the only recent exception I know about, see Tegmark (1996c).] Odious as this practice is, I have bowed to convention in these lectures.

3.3. ACOUSTIC OSCILLATIONS

To develop an intuitive feel for the solutions to equation (12), we will start by making some excessive and unwarranted approximations. We will then gradually relax those approximations to get a more accurate picture. First, let's assume that R and Ψ are independent of time. Then

$$(1+R)\ddot{\Theta} + \frac{k^2}{3}\Theta = -\frac{k^2}{3}(1+R)\Psi. \tag{18}$$

This is the equation for a simple harmonic oscillator, with solution

$$\Theta(\eta) = -(1+R)\Psi + K_1\cos(kc_s\eta) + K_2\sin(kc_s\eta). \tag{19}$$

Here K_1 and K_2 are constants to be fixed by the initial conditions and $c_s = (3(1+R))^{-1/2}$ is the sound speed. In this approximation, then, each Fourier mode represents an acoustic plane wave propagating at speed c_s.

There is a simple physical picture underlying this result. The baryon-photon fluid wants to fall into the potential wells, but it is supported by radiation pressure. The balance between pressure and gravity sets up acoustic oscillations. The three terms in equation (12) come from the inertia of the fluid, the radiation pressure, and the gravitational field.

In fact, let's make things even simpler and set $R = 0$. Then

$$\Theta(\eta) = -\Psi + K_1\cos(kc_s\eta) + K_2\sin(kc_s\eta). \tag{20}$$

In many theories, the initial perturbation is *adiabatic*, meaning that the matter and radiation fluctuations are the same at any particular point. With these initial conditions, $\dot{\Theta} = 0$ at very early times, and $\Theta(0) = -2\Psi/3$, so

$$\Theta(\eta) = -\Psi + \tfrac{1}{3}\Psi\cos kc_s\eta. \tag{21}$$

Continuing to focus our attention on a single Fourier mode, let us determine what kind of anisotropy we would expect to see on the sky. As we have mentioned, the three sources of primary anisotropy are gravity, the Doppler effect, and intrinsic temperature variations,

$$\frac{\Delta T}{T} = \left[\Psi + \hat{\mathbf{r}}\cdot\mathbf{v} + \Theta\right]_{\eta=\eta_{\rm LS}}, \tag{22}$$

where $\eta_{\rm LS}$ is the time of last scattering and $\hat{\mathbf{r}}$ is a unit vector in the direction of observation.

Ignoring the Doppler term for the moment, note that the other two terms give a pure cosine oscillation,

$$\Psi + \Theta = \tfrac{1}{3}\Psi\cos kc_s\eta, \tag{23}$$

so the r.m.s. $\Delta T/T$ is large when $kc_s\eta_{\rm LS}$ is an integer multiple of π. Therefore, if the initial conditions have a smooth power spectrum, $\Delta T/T$ will have a harmonic series of peaks in k-space, leading to a harmonic series in the angular power spectrum of anisotropy on the sky. This is the origin of the so-called "Doppler peaks" in Figure 1. Ironically, the peaks have nothing to do with the Doppler effect. In fact, the peaks are caused by modes that have reached maxima of compression and rarefaction at the time of last scattering; the Doppler contribution to the anisotropy in these modes is zero!

The first peak is caused by modes that have had time to oscillate through exactly one half of a period before last scattering; the modes that cause the second peak have oscillated through a full period, and so on. The physical scale of the first peak is therefore $\lambda \sim k^{-1} = c_s\eta_{\rm LS}/\pi \sim 30\,{\rm Mpc}$. The distance to the last-scattering surface is $D \equiv \eta_0 - \eta_{\rm LS} \sim 6000\,{\rm Mpc}$, so the angular scale of the first peak is $\lambda/D \sim 0.^\circ25$. We will be more precise about the correspondence between physical scales and angular scales later.

Earlier, we threw out the Doppler term in equation (22) for no particular reason. We had better put it back. Using the continuity equation (1) and the relation $\delta = 3\Theta$, we find that

$$\mathbf{v} = \frac{3i}{k}\dot{\Theta}\hat{\mathbf{k}}. \tag{24}$$

Here $\hat{\mathbf{k}}$ is a unit vector in the direction of $\mathbf{k}$ and $\mathbf{v}$ and δ are still in Fourier space. Differentiating equation (21) and using the fact that $c_s = 1/\sqrt{3}$ for $R = 0$, we find that $\dot{\Theta} = -\frac{1}{3\sqrt{3}}\Psi \sin kc_s\eta$. Since the r.m.s. value of $\hat{\mathbf{r}}\cdot\hat{\mathbf{k}}$ is $1/\sqrt{3}$, the r.m.s. Doppler contribution to equation (22) is

$$\left[\frac{\Delta T}{T}\right]_{\rm Doppler} = \frac{i}{3}\Psi \sin kc_s\eta. \tag{25}$$

This has the same amplitude as the $(\Theta + \Psi)$ contribution, but is 90° out of phase in both time (it goes like a sine instead of a cosine) and space (it has an extra factor of i). This has the rather disastrous consequence of completely erasing the Doppler peaks: the total $\Delta T/T$ is the quadrature sum of (23) and (25):

$$\left(\frac{\Delta T}{T}\right)^2 \propto \sin^2 kc_s\eta + \cos^2 kc_s\eta = 1. \tag{26}$$

The k dependence, which led to the peaks, is gone.

The problem, of course, is that we have taken our approximations too far. Specifically, the culprit is the limit $R \to 0$. Physically, taking the limit

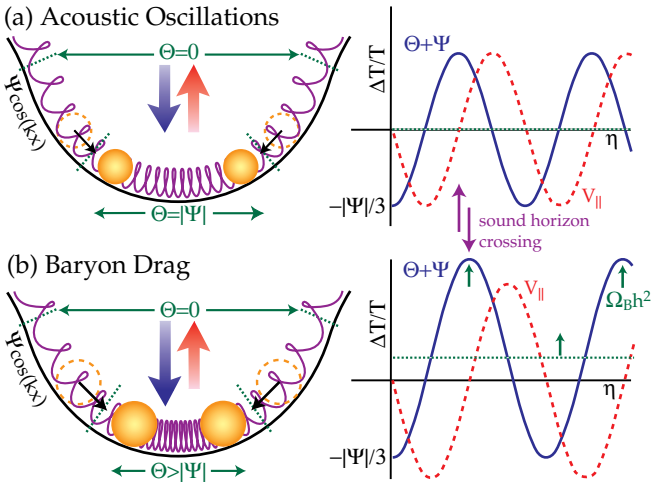

Figure 2. A simple mechanical model for a single mode of acoustic oscillation of the photon-baryon fluid. The behavior of the fluid inside of a potential well is shown; the behavior atop a potential hill would be the reverse. The springs represent the restoring force of the photon pressure and the balls represent the effective mass of the system. The top panel shows the case where the baryon contribution to the effective mass can be neglected, and the lower panel shows the effect of including baryons. Baryons increase the mass of the fluid, causing a displacement of the zero point of the oscillations. In addition, the sound speed is lowered. This has two effects, both of which may be seen in the plots on the right: baryons make the oscillations proceed more slowly and also reduce the Doppler contribution to $\Delta T/T$ relative to the intrinsic and Sachs-Wolfe contributions. Reprinted from Hu (1996).

$R \to 0$ means ignoring the dynamical effects of the baryons. Let us remove that assumption, but keep the approximation that R is time-independent. Then the solution for $\Theta(\eta)$ changes in two ways. The sound speed gets smaller by a factor $1/\sqrt{1+R}$, and the driving term $F(\eta)$ gets bigger by a factor $1+R$. The adiabatic solution to equation (12) is now

$$\Theta(\eta) = \tfrac{1}{3}(1+3R)\Psi \cos kc_s\eta - (1+R)\Psi. \tag{27}$$

By allowing R to be nonzero, we have increased the amplitude of the cosine oscillations by a factor $(1+3R)$. Furthermore, there is now an offset in the combined Sachs-Wolfe and adiabatic contributions to $\Delta T/T$: in the limit $R \to 0$, we found that $\Theta + \Psi$ oscillated symmetrically about zero; now it oscillates about $-R\Psi$. Most important, a nonzero R reduces the

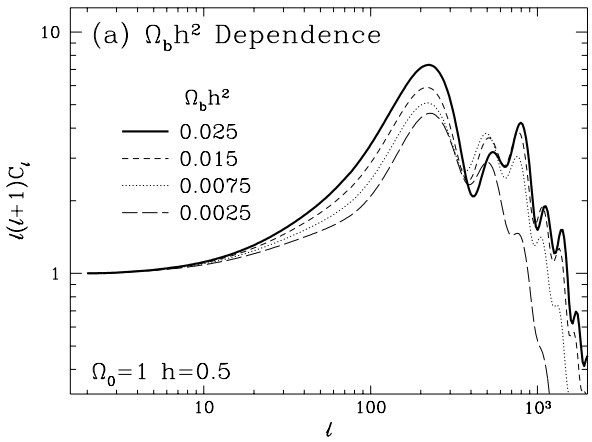

Figure 3. Angular power spectra for CDM models with varying values of the baryon density $\Omega_B h^2$. Reprinted from Hu (1996).

amplitude of the Doppler contribution to the anisotropy, relative to the Sachs-Wolfe contribution, since v is proportional to $c_s\Theta$ and c_s has gotten smaller. Since the cosine oscillations are now larger in amplitude than the sine oscillations, we do indeed expect to see a series of peaks at $kc_s\eta_{\rm LS} = m\pi$.

Why does including the dynamical effect of the baryons effect these changes in the solution? The essential reason is that baryons contribute to the *effective mass* of the photon-baryon fluid, but not to the *pressure*. (This is clear from looking at equation (12): the first term, representing the effective mass, depends on R, but the second term, representing pressure support, does not.) The effect of the baryons, therefore, is to slow down the oscillations, and also to make the fluid fall deeper into the potential wells. This explains all three of the key effects we have just mentioned: the increased oscillation amplitude, the offset in the center of the oscillations, and the reduction in importance of the velocity term relative to the other terms. These effects are represented pictorially in Figure 2.

Based on this analysis, we can predict that the height of the peaks in the CMB anisotropy spectrum should depend on the baryon density: the larger the baryon density, the larger R, and the greater the amplitude of the oscillations. Furthermore, because of the offset in the oscillations, we expect the odd-numbered peaks to be enhanced relative to the even-numbered ones. (In the language of Figure 2, the compressions produce larger anisotropies than the rarefactions. Of course, if we had chosen to draw a potential peak instead of a potential well in Figure 2, we would

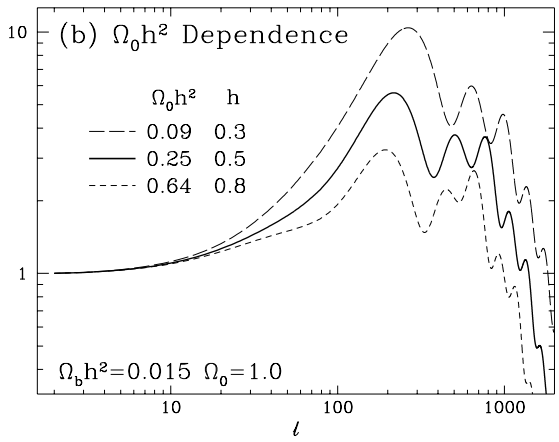

Figure 4. Angular power spectra for CDM models with varying values of h. All of these models have $\Omega_0 = 1$. For lower values of $\Omega_0 h^2$, matter domination occurs later. The driving effect of the decay in the gravitational potential is therefore more significant, increasing the peak height. Reprinted from Hu (1996).

make precisely the opposite statement.)

Both of these effects are found in detailed calculations and can be seen in Figure 3.

We can make further refinements to these approximations without too much difficulty. For instance, we can allow R to vary with time. The time scale on which R varies is of order a Hubble time and is much longer than the period of the acoustic waves. We can therefore treat the variation of R (and the concomitant variation in c_s) in the WKB approximation. There are two main results. First, the phase of the oscillation changes from $kc_s\eta$ to $k\int c_s\, d\eta$. Second, the amplitude of the oscillations grows with time in proportion to $c_s^{1/2}$, or $(1+R)^{-1/4}$.[6]

3.4. DRIVING

We can also relax the approximation that $F(\eta)$ is constant in time. This has interesting consequences. A constant term on the right-hand side of an oscillator equation merely offsets the center of the oscillations; in contrast, a time-varying term genuinely drives oscillations. In particular, if the driving

[6]The easiest way to see this is to note that $m\omega A^2$ is an adiabatic invariant for a harmonic oscillator. Here m is the mass, A is the amplitude, and ω is the frequency. Of course, the result can also be derived directly from the WKB approximation.

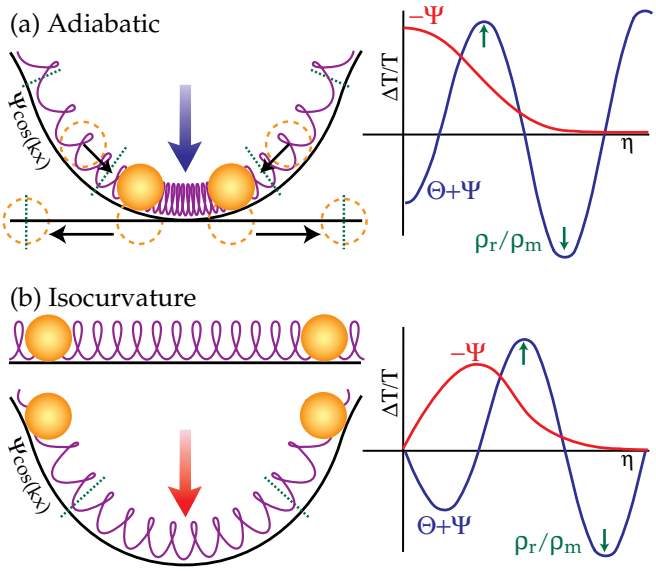

Figure 5. Driving effects on the acoustic oscillations. (a) In the adiabatic case, the gradual decay in the potential causes a relatively small increase in the amplitude of the oscillations. (b) For isocurvature initial conditions, the initial perturbation $\Theta + \Psi$ is zero, and the growth (and subsequent decay) of Ψ is entirely responsible for driving the oscillations. Reprinted from Hu, Sugiyama, & Silk (1996).

term varies significantly on a time scale comparable to the period of the oscillations, resonant driving can occur.

We have seen that Ψ (and hence F) is constant during matter domination, but it decays during the radiation epoch. For modes that enter the horizon before matter domination, Ψ decays while that mode is undergoing its oscillations. The decay in Ψ therefore boosts the amplitude of those short-wavelength modes. The modes that receive the largest boost are those that entered the horizon before matter-radiation equality at a redshift

$$z_{\rm eq} = 24000\Omega_0 h^2. \tag{28}$$

These modes are characterized by wavenumbers

$$k \gtrsim k_{\rm eq} = (14\,{\rm Mpc})^{-1}\Omega_0 h^2. \tag{29}$$

The effect of the driving term becomes evident if we look at power spectra for critical-density models with different values of the Hubble parameter: for low h, matter domination occurs later and the boosting effect is greater. This effect is shown in Figure 4.

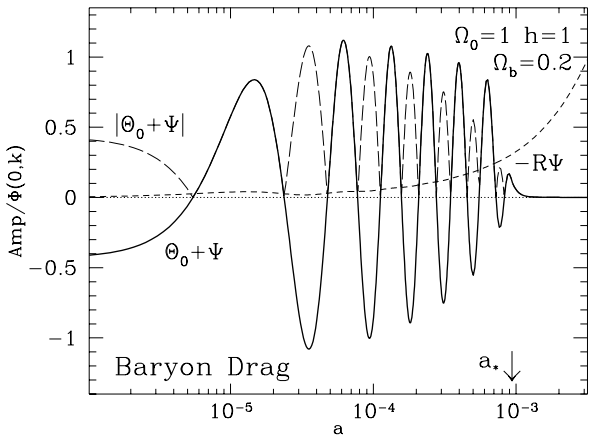

Figure 6. The time evolution of a single Fourier mode. a is the scale factor, normalized to unity today. a_* is the scale factor at recombination. At early times potential decay increases the amplitude of the oscillations. The heights of the positive and negative peaks are offset by $-R\Psi$ with respect to each other. The decline in amplitude at late times is due to diffusion damping. Reprinted from Hu (1996).

We have been focusing on models with adiabatic initial conditions. If we instead consider *isocurvature* models, the effect of the driving term becomes even more evident. In isocurvature models, the total density perturbation vanishes at early times:

$$\delta\rho_{\rm total} = \delta\rho_{\rm B} + \delta\rho_\gamma + \delta\rho_{\rm CDM} + \ldots = 0. \tag{30}$$

Clearly $\Theta(0) = 0$ in these models. As time passes, $\delta\rho_\gamma$ redshifts away, leaving genuine density perturbations and hence nonzero potentials Φ and Ψ. Oscillations are therefore driven in Θ. In contrast to the adiabatic case, these isocurvature oscillations are proportional to $\sin kc_s\eta$ rather than $\cos kc_s\eta$. The peaks in an isocurvature spectrum are therefore different in phase from adiabatic peaks. The peak locations in the CMB anisotropy spectrum can distinguish quite robustly between adiabatic and isocurvature models. Figure 5 illustrates the origin of the peaks in isocurvature models.

3.5. DAMPING

We have been assuming so far that the tight-coupling approximation holds perfectly right up until the moment $\eta_{\rm LS}$, and that the photons are instantaneously released at that moment. In fact, the failure of the tight-coupling approximation, especially around the time of last scattering, causes sig-

nificant damping of fluctuations as photons diffuse out of hot, overdense regions. Furthermore, the last-scattering "surface" is really a shell of some thickness. Oscillations on scales smaller than this thickness do not show up as observable anisotropies on the sky, since any particular line of sight will look at multiple peaks and troughs of that mode.

To get a rough estimate of the importance of diffusion damping (also known as Silk damping), consider a photon undergoing a random walk through the photon-baryon fluid. If the mean free path is λ, then at a time η, a typical photon has scattered about $N \sim \eta/\lambda$ times and has diffused through a distance $\lambda_{\rm D} \sim \sqrt{N}\lambda \sim \sqrt{\eta\lambda}$. If a particular Fourier mode has a wavelength less than this diffusion length, then the photons will have diffused from overdense to underdense regions, and the mode will be damped away. Diffusion damping thus occurs for modes with $k^{-1} \gtrsim \lambda_{\rm D}$. Most of the damping occurs around the time of last scattering, since that is when the mean free path λ becomes large.

In Figure 6 we show the time evolution of a particular mode, including the damping at the end, and in Figure 10 below we show the net effect of diffusion damping on a CMB power spectrum.

3.6. PROJECTION

In order to complete the story of primary anisotropies, we need to specify precisely how a particular plane wave is projected onto a specific angular scale on the sky. It is clear that a mode with wavelength λ will show up on an angular scale $\theta \sim \lambda/R$, where R is the distance to the last-scattering surface, or in other words, a mode with wavenumber k shows up at multipoles $l \approx k$. Consequently, tilting the spectral index n of the primordial matter power spectrum essentially just tilts the angular power spectrum. Let us now make this rough observation mathematically precise.

If we are looking in a direction $\hat{\mathbf{r}}$ in the sky, then (ignoring the thickness of the last-scattering surface), the anisotropy we see is simply $\Delta T/T(\hat{\mathbf{r}}) = \Theta^{(\rm tot)}(R\hat{\mathbf{r}})$, where $\Theta^{(\rm tot)}$ includes all three terms in equation (22). For a single Fourier mode, this is simply

$$\frac{\Delta T}{T}(\hat{\mathbf{r}}) = \Theta^{(\rm tot)}_{\mathbf{k}} \exp(i\mathbf{k}\cdot\hat{\mathbf{r}}R). \tag{31}$$

To quantify the amount of power this produces on different angular scales, we expand in spherical harmonics Y_{lm}. The relevant identity is (Jackson 1975)

$$\exp(ikR\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}) = 4\pi\sum_{l,m} i^l j_l(kR) Y^*_{lm}(\hat{\mathbf{k}}) Y_{lm}(\hat{\mathbf{r}}). \tag{32}$$

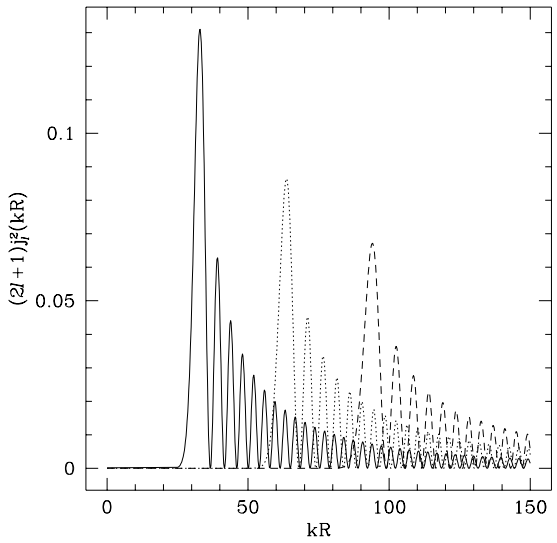

Figure 7. The quantity $(2l+1)j_l^2(kR)$ is plotted for $l = 30$, $l = 60$, and $l = 90$. This quantity determines how much power a Fourier mode with wavenumber k contributes to multipole l. Note that, while most of the power is deposited at $l \simeq kR$, there is significant "bleeding" to lower l.

Combining equations (31) and (32), we find that

$$\frac{\Delta T}{T}(\hat{\mathbf{r}}) = \sum_{l,m} a_{lm} Y_{lm}(\hat{\mathbf{r}}), \tag{33}$$

where

$$a_{lm} = 4\pi \Theta_{\mathbf{k}}^{(\mathrm{tot})} i^l j_l(kR) Y_{lm}^*(\hat{\mathbf{k}}). \tag{34}$$

The total power produced by this mode in the multipole l is

$$a_l^2 \equiv \sum_{m=-l}^{l} |a_{lm}|^2 = 4\pi(2l+1) \left|\Theta_{\mathbf{k}}^{(\mathrm{tot})}\right|^2 j_l^2(kR). \tag{35}$$

The spherical Bessel function $j_l(x)$ peaks at $x \sim l$, so a single Fourier mode $\mathbf{k}$ does indeed contribute most of its power around multipole $l_k = kR$, as expected. However, as Figure 7 shows, j_l does have significant power beyond the first peak, meaning that the power contributed by a Fourier mode "bleeds" to l-values lower than l_k. This is due to the fact that a mode appears to have a longer wavelength when looked at along a line of sight nearly perpendicular to the wavevector.

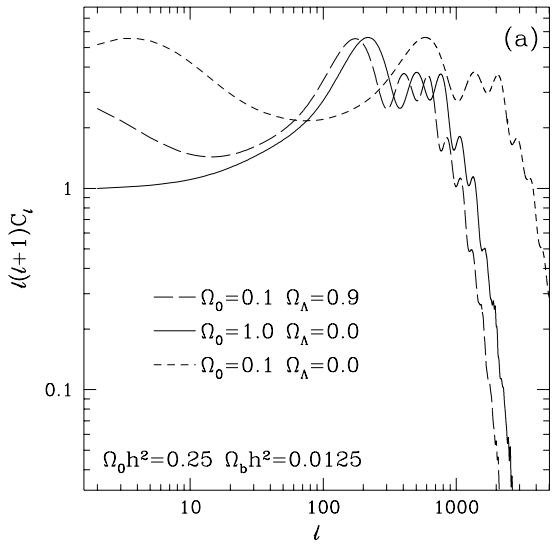

Figure 8. Ω_0-dependence of the angular power spectrum. In open models, the angular-diameter distance to the last-scattering surface is large, so the features in the power spectrum are shifted to small angular scales. In a flat model with a cosmological constant, the distance to the last-scattering surface is larger than in an $\Omega_0 = 1$ model, but the size of the sound horizon also increases, producing little net effect on the location of the peaks. The structure at low l in the low-density models is due to the integrated Sachs-Wolfe effect. Reprinted from Hu & White (1996).

These formulae assume that the Universe is spatially flat. If there is curvature, then the correspondence between physical scales at last scattering and angular scales on the sky changes. In an open Universe, for example, geodesics focus in such a way that a particular angular scale corresponds to a much larger physical scale on the last-scattering surface. A particular Fourier mode in an open Universe projects to multipoles $l \sim kR_A$, where R_A is the *angular-diameter distance* to the last-scattering surface, given by

$$R_A = \frac{1}{\sqrt{|K|}} \sinh\left(\sqrt{|K|}R\right). \tag{36}$$

Here K is the curvature. When $|K|$ is small, $R_A \to R$, but for large $|K|$ (low Ω_0), R_A grows exponentially with metric distance.

This projection effect is easy to see in predictions of the CMB anisotropy. In an open Universe, features such as the acoustic peaks and the damping scale are shifted towards smaller angular scales, *i.e.*, towards higher l. (See Figure 8.)

Note that the approximate linear relation between l and k holds only for primary anisotropies. The secondary anisotropies, which we discuss below,

tend to occur at a wide range of distances (in contrast to the relatively thin last-scattering surface). Thus for secondary anisotropies, each k-mode can contribute to a wide range of l's.

4. Secondary Anisotropies

After last scattering, the photons and baryons are no longer tightly coupled. In fact, if the effects of reionization are negligible, there is no coupling at all. In this case, the photons simply propagate freely along spacetime geodesics from last scattering to the observer. The causes of secondary anisotropy are then entirely gravitational, the dominant effect being the ISW effect. Weak gravitational lensing can also distort the anisotropy spectrum, although this effect is generally small.

If the intergalactic medium reionized at a sufficiently early redshift, then some fraction of the photons will interact again after the time of "last" scattering. The main result is that primary fluctuations are erased, and in addition new fluctuations can be generated from the new last-scattering surface. However, the last-scattering surface in a reionized model is extremely thick (since the photon-baryon coupling is weak), so the nature of the regenerated anisotropy is quite different from the primary anisotropy.

4.1. INTEGRATED SACHS-WOLFE EFFECT

As Sachs & Wolfe (1967) showed, fluctuations in the spacetime curvature produce CMB anisotropy in two distinct ways. The "ordinary" Sachs-Wolfe effect is simply the gravitational red- or blueshift due to the potential difference between the points of emission and reception of a photon. In addition, if the gravitational potential changes with time, there is an "integrated" Sachs-Wolfe effect.

Imagine a photon falling into a potential well, and then climbing out the other side. If the potential does not vary with time, the photon suffers no net change in energy. However, if the potential well decays while the photon is passing through it, then the redshift upon climbing out of the well is smaller than the blueshift upon falling in. The photon therefore gains energy. The magnitude of the ISW effect is given by an integral along the photon's path:

$$\left(\frac{\Delta T}{T}\right)_{ISW} = \int \left(\dot{\Psi}(\mathbf{x},\eta) - \dot{\Phi}(\mathbf{x},\eta)\right)\, d\eta. \tag{37}$$

We observed earlier that the gravitational potential is time-independent if certain conditions are satisfied:

- The Universe is matter-dominated ($\rho_{\rm matter} \gg \rho_{\rm rad}$).

- Spatial curvature is negligible ($\Omega = 1$).
- Linear perturbation theory is valid ($\delta \ll 1$).

If all of these conditions are satisfied, there is no ISW effect. However, in any realistic cosmological model some or all of these conditions are violated at some point.

4.2. EARLY ISW EFFECT

In a typical model, the epoch of matter-radiation equality occurs before the time of last scattering, but not long before. The matter-dominated limit is therefore not quite correct around the time of last scattering and shortly thereafter. The decay in the potential shortly after last scattering gives rise to the early ISW effect. This effect is largest when the matter density $\Omega_0 h^2$ is low.

The early ISW effect is most important on large scales. Specifically, the scales that are most affected are those with k^{-1} comparable to the time scale on which the potential decays. Modes with wavelengths much shorter than this oscillate many times while the potential is decaying, causing both positive and negative ISW contributions, which tend to cancel each other out. The time scale for potential decay is of order the horizon size at last scattering, so the early ISW effect shows up on large angular scales $l \lesssim 200$.

4.3. LATE ISW EFFECT

In models with $\Omega_0 \neq 1$, the potential decays at late times, typically at redshifts $z \lesssim \Omega_0^{-1}$. This potential decay, which occurs whether or not there is a cosmological constant, gives rise to an ISW effect at late times. As with the early ISW effect, modes with wavelengths comparable to the time scale for the potential to decay are most affected. The relevant time scale is the horizon size at the time of potential decay, so the late ISW effect also leaves its imprint on large angular scales.

4.4. OTHER ISW EFFECTS

At very late times, nonlinear structure forms, causing the potential to grow with time. The ISW effect due to nonlinear structure is often called the *Rees-Sciama effect* (Rees & Sciama 1968). In standard models, the Rees-Sciama effect is typically much weaker than the other effects we have discussed (Seljak 1996a).

A background of primordial gravity waves, if there is one, produces its own ISW effect. Gravity waves redshift once they enter the horizon, so modes that enter the horizon well before last scattering leave no imprint on the CMB. The gravity-wave contribution to the CMB anisotropy therefore

occurs on large angular scales $l \lesssim 100$. Because of the quadrupolar nature of the spacetime distortion caused by a gravity wave, the gravity-wave contribution to the CMB quadrupole is enhanced relative to other modes.

There may be other sources of spacetime distortion besides linear density fluctuations and gravity waves. In particular, topological defects cause spacetime curvature and hence an ISW effect. We will not discuss topological defects further; for more information, see Paul Shellard's contribution to this volume, and the references therein.

4.5. GRAVITATIONAL LENSING

The ISW effect may be thought of as gravity imparting a "kick" to a photon forward or backward along the direction of motion. Gravity can also kick the photons in the transverse directions, changing their directions of motion but not their energies. The result of this weak gravitational lensing is that our image of the last-scattering surface is slightly distorted, as if we were looking at it through an irregular refracting medium. This distortion of the last-scattering surface results in a slight smearing of the angular power spectrum, with power from the peaks being moved into the valleys. The effect is typically weak, resulting in changes at the few-percent level in the power spectrum (Seljak 1996b).

4.6. REIONIZATION

We will not undertake a detailed discussion of reionized models here. Instead, we refer the interested reader to Roman Juszkiewicz's contribution to this volume and references therein. We will, however, make some general comments.

The Gunn-Peterson test (Gunn & Peterson 1965) tells us that the intergalactic medium is ionized out to redshifts of a few. In CDM-like models of structure formation, reionization is generally thought to occur at such moderate redshifts, with the formation of the earliest nonlinear structures. If this is correct, then reionization does not dramatically alter the CMB anisotropy predictions. If, on the other hand, reionization somehow happened earlier, say at $z \gtrsim 100$, then a significant fraction of the CMB photons have been scattered by the reionized matter after the so-called epoch of last scattering.

The main effect of such early reionization is to erase anisotropy on degree scales. The reason is quite simple: if we have early reionization, then a photon that comes toward us from a particular direction need not have originated from that direction. Rather, as Figure 9 illustrates, each direction on the sky contains photons that originate from a variety of different locations at the time of "last" scattering. In severely reionized models, the

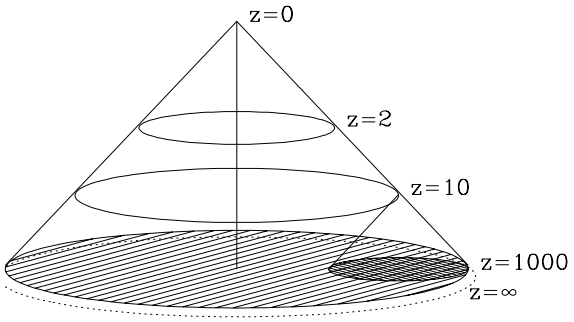

Figure 9. Our backward light cone. The vertical axis represents conformal time, and the horizontal axes are two of the three spatial directions. In the absence of reionization, each line of sight corresponds to a particular point on the last-scattering surface at $z \simeq 1000$. In a reionized model in which a typical photon last scattered at $z = 10$, a photon arriving from a particular direction may have originated from any point in the shaded circle. Reprinted from Tegmark (1996c).

peaks are completely washed away. Such models may already be ruled out by degree-scale CMB experiments (Scott, Silk, & White 1995).

Inhomogeneities and bulk motions of the reionized matter induce new CMB anisotropies, which must generally be treated to second order in perturbation theory (Ostriker & Vishniac 1986; Hu, Scott, & Silk 1994; Dodelson & Jubas 1995), but we will not discuss these regenerated anisotropies here. We also neglect to discuss the effect of nonuniform or patchy reionization, including the Sunyaev-Zel'dovich effect (Sunyaev & Zel'dovich 1970).

5. Summary of Anisotropy Formation

We have now concluded our tour of the mechanisms of anisotropy formation. Figure 10 illustrates some of the key points. The dominant features in a typical CDM power spectrum are the peaks due to acoustic oscillations of the photon-baryon fluid. The peaks correspond to modes that are undergoing maximum compression and rarefaction at the time of last scattering. Modes that are out of phase with these modes produce anisotropy via the Doppler effect, partially filling in the valleys between the acoustic peaks. The effect of damping on small scales is evident, and the rise at $l \sim 500$ in the undamped spectrum shows the driving effect of the decaying gravitational potential at early times.

We can use what we have learned to determine how the predicted anisotropy spectrum should depend on the key cosmological parameters:

- In models with *spatial curvature* ($\Omega_0 \neq 1$), the position of the acoustic

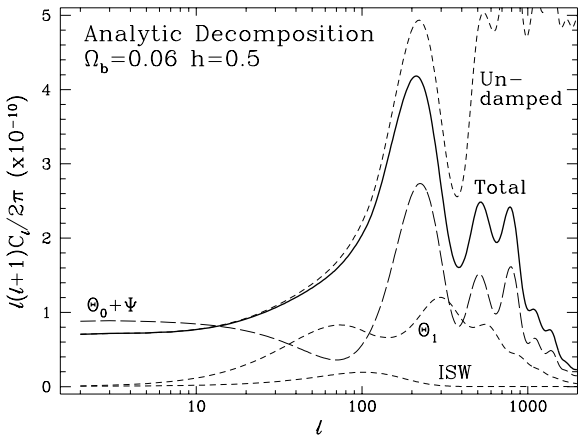

Figure 10. Analytic decomposition of anisotropies. The solid line shows the angular power spectrum of a critical-density CDM model. The upper dashed curve shows the spectrum that would be seen in the absence of diffusion damping. Note that the undamped peak heights increase at scales small enough to have crossed the horizon before matter-radiation equality. (See Section 3.4.) The other curves show the relative importance of the Sachs-Wolfe, integrated Sachs-Wolfe, and Doppler (Θ_1) contributions. Reprinted from Hu (1996).

peaks shifts due to geodesic deviation. In addition, the late ISW effect boosts the large-scale power.

- If there is a *cosmological constant*, then the position of the peaks shifts slightly due to the increased distance to the last-scattering surface, and again, the late ISW effect boosts the large-scale power.
- Lowering the *Hubble parameter* (for fixed Ω_0) reduces the matter density. The gravitational driving of oscillations is enhanced, and the peaks increase in height.
- The higher the *baryon density* $\Omega_{\rm B}h^2$, the greater the peak amplitude. Odd-numbered peaks in particular are enhanced.
- The *spectral index* n of the primordial power spectrum essentially just tilts the angular power spectrum.
- If we add *gravity waves* to a model, we increase the quadrupole, and in addition the whole "plateau" at low l rises relative to the acoustic peaks.

Although we have made many approximations in deriving these conclusions, all of them are borne out by detailed Boltzmann calculations. Because the CMB anisotropy predictions depend sensitively on the various parameters, an experiment that could map out the acoustic peaks would be able to measure these cosmological parameters accurately. The spatial

curvature in particular should be relatively easy to pick out, thanks to the shift in position of the first peak. The relative positions of successive peaks also provide a robust way of determining whether the initial conditions are isocurvature or adiabatic.[7] The dependence of the power spectrum on other parameters such as h and $\Omega_{\rm B}$ is somewhat more subtle, but if we manage to detect and measure the heights of two or three peaks, we should be able to do quite well (Jungman *et al.* 1996), assuming, of course, that the general paradigm sketched above is correct and the multiple peaks are really there.

6. Statistical Properties of $\Delta T/T$

Before we discuss methods for comparing theories with data, we need to discuss briefly the statistical properties of the CMB anisotropy as it appears on the sky. As we have mentioned, it is convenient to expand the observed anisotropy in spherical harmonics:

$$\frac{\Delta T}{T}(\hat{\mathbf{r}}) = \sum_{l,m} a_{lm} Y_{lm}(\hat{\mathbf{r}}). \tag{38}$$

We have focused on the anisotropy produced by an individual plane wave; the observed anisotropy is of course a superposition of contributions from all of these plane waves:

$$a_{lm} = \sum_{\mathbf{k}} a_{lm}^{(\mathbf{k})}. \tag{39}$$

Since all of the relevant physics is described by linear perturbation theory — as we know, everything in nature is linear[8] — each $a_{lm}^{(\mathbf{k})}$ is proportional to the initial density perturbation $\delta_{\mathbf{k}}^{(\rm init)}$.

One often assumes that the initial conditions have "random phases," meaning that different Fourier modes are uncorrelated,

$$\langle \delta_{\mathbf{k}} \delta_{\mathbf{k}'} \rangle = 0 \qquad \text{when} \qquad \mathbf{k} \neq \mathbf{k}'. \tag{40}$$

In this case, the $a_{lm}^{(\mathbf{k})}$ are also uncorrelated. The mean-square power in a particular multipole is then simply the sum of the contributions from the various Fourier modes:

$$\langle |a_{lm}|^2 \rangle = \sum_{\mathbf{k}} \langle |a_{lm}^{(\mathbf{k})}|^2 \rangle. \tag{41}$$

[7] For isocurvature initial conditions, the second and third peaks occur at three and five times the location of the first; in adiabatic models, they occur at two and three times.

[8] to first order.

And, of course, the left-hand side of this equation is simply the angular power spectrum C_l. (This quantity is independent of the azimuthal index m as long as space is isotropic.)

We often go beyond the assumption of random phases and assume *Gaussian initial conditions*. This is a prediction of inflationary scenarios, but one often assumes Gaussian initial conditions even in non-inflationary phenomenological models such as isocurvature baryon models (Peebles 1987). When we talk about a Gaussian theory, we simply mean that at some initial time t_i the density perturbation δ was a realization of a Gaussian random field. Bernard Jones has provided a detailed discussion of Gaussian random processes elsewhere in this volume; for our purposes, all we need to know is that the assumption of Gaussian initial conditions, together with homogeneity and isotropy, implies that each Fourier coefficient $\delta_{\mathbf{k}}$ is an independent Gaussian random variable of zero mean.[9] In other words, the real-space density perturbation $\delta(\mathbf{x})$ is a stochastic superposition of plane waves of all different wavelengths. Since a Gaussian random variable is completely determined by its mean and variance, and since $\langle \delta_{\mathbf{k}} \rangle = 0$, the statistical properties of our Gaussian random field are completely determined by the power spectrum $P(k) \equiv \langle |\delta_{\mathbf{k}}|^2 \rangle$.

If we assume Gaussian initial conditions, then each coefficient a_{lm} is a Gaussian random variable, since it is a linear combination of the Gaussian variables $\delta_{\mathbf{k}}$. The statistical properties of $\Delta T/T$ are therefore completely specified by the means,

$$\langle a_{lm} \rangle = 0, \tag{42}$$

and the covariances,

$$\langle a_{lm} a^*_{l'm'} \rangle = C_l \delta_{ll'} \delta_{mm'}, \tag{43}$$

of the coefficients a_{lm}. In other words, for Gaussian initial conditions, the angular power spectrum C_l tells us everything we need to know.

Even when the initial conditions are not Gaussian, it often suffices to treat the CMB anisotropy as Gaussian, at least on sufficiently large angular scales. The CMB fluctuation on large angular scales is typically due to a superposition of many incoherent fluctuations. Even if the individual fluctuations fail to be Gaussian, the central limit theorem guarantees that the superposition will be approximately Gaussian. When comparing the COBE data with the predictions of a cosmic string model, for example, it is perfectly adequate to treat $\Delta T/T$ as Gaussian, even though the underlying perturbations are highly non-Gaussian.

[9]That is, Gaussian initial conditions (together with homogeneity and isotropy) imply random phases, but not conversely.

7. An Introduction to CMB Data Analysis

7.1. AN IDEALIZED EXPERIMENT

We will explore the key issues in CMB data analysis by first considering an absurdly idealized experiment (the sort of thing only a theorist could dream up). We will gradually introduce real-world complications to see what the main issues are.

Imagine, then, an experiment that measured $\Delta T/T$ at many pixels that cover the entire sky completely and uniformly. Furthermore, imagine that each data point is a perfect, noise-free measurement. With this data set, we could determine each coefficient a_{lm} with essentially perfect accuracy by inverting equation (38):

$$a_{lm} = \int \frac{\Delta T}{T}(\hat{\mathbf{r}}) Y^*_{lm}(\hat{\mathbf{r}})\, d\Omega \approx \frac{4\pi}{N_{\rm pix}} \sum_{p=1}^{N_{\rm pix}} \frac{\Delta T}{T}(\hat{\mathbf{r}}_p) Y^*_{lm}(\hat{\mathbf{r}}_p). \tag{44}$$

Here $d\Omega$ is an element of solid angle in the direction of $\hat{\mathbf{r}}$, $N_{\rm pix}$ is the total number of pixels, and $\hat{\mathbf{r}}_p$ is a unit vector in the direction of the pth pixel.

Even in this hopelessly idealized experiment, we still can't measure the angular power spectrum C_l perfectly. The reason is that C_l is an ensemble-average quantity: it is the variance of the distribution from which a_{lm} is drawn. We have only a finite number, $2l+1$, samples of this distribution at each l. This fact, generally called *cosmic variance*, sets a fundamental limit on how well we can ever hope to measure the angular power spectrum.[10]

If we assume Gaussian statistics, then the best estimator of C_l is simply the average of $|a_{lm}|^2$ over m:

$$\hat{C}_l \equiv \frac{1}{2l+1} \sum_{m=-l}^{l} |a_{lm}|^2. \tag{45}$$

This quantity is chi-squared distributed with $2l+1$ degrees of freedom, and so it has a fractional uncertainty of

$$\frac{{\rm Var}^{1/2}(\hat{C}_l)}{C_l} = \sqrt{\frac{2}{2l+1}}. \tag{46}$$

The unfortunate fact, therefore, is that even in a perfect experiment we will never know C_l with a fractional uncertainty better than $(l+\frac{1}{2})^{-1/2}$. We are stuck with a 63% uncertainty in the quadrupole power C_2 and a 30%

[10]Cosmic variance is closely related to the failure of ergodicity. If $\Delta T/T$ were ergodic, then the average value of $|a_{lm}|^2$, measured in different orientations over the sphere, would be the ensemble-average quantity C_l. But $\Delta T/T$ isn't ergodic, so this doesn't work.

uncertainty in C_{10}, although we can in principle hope to determine C_{1000} to 0.3%.

7.2. NOISE

Let's mess up our nice, clean experiment by adding noise. Each pixel is no longer a perfect measurement of $\Delta T/T$: the ith data point d_i consists of a sum of signal and noise,

$$d_i = \frac{\Delta T}{T}(\hat{\mathbf{r}}_i) + n_i. \tag{47}$$

Let us assume that the noise n_i in each pixel is independent and Gaussian distributed, with some standard deviation σ. For the moment we will assume *homoskedasticity*, that is, that σ is the same in all pixels.

We can still try to estimate a_{lm} using equation (44),

$$\hat{a}_{lm} = \frac{4\pi}{N_{\rm pix}} \sum_{p=1}^{N_{\rm pix}} d_p Y^*_{lm}(\hat{\mathbf{r}}_p), \tag{48}$$

and average over m to get an estimate of C_l,

$$\hat{C}_l = \frac{1}{2l+1} \sum_{m=-l}^{l} |\hat{a}_{lm}|^2, \tag{49}$$

but this quantity will no longer be a good estimate of the true C_l; it will be biased upward. Using equations (48) and (47), together with the fact that $\langle n_p n_{p'} \rangle = \sigma^2 \delta_{pp'}$, it is straightforward to check that

$$\langle |\hat{a}_{lm}|^2 \rangle = C_l + \frac{4\pi}{N_{\rm pix}} \sigma^2. \tag{50}$$

The estimator $\hat{C}_l$ is the average of these quantities, so it too is biased upward by $4\pi\sigma^2/N_{\rm pix}$.

We can of course get a better estimate of C_l by subtracting off the noise bias,

$$\hat{C}'_l \equiv \hat{C}_l - \frac{4\pi}{N_{\rm pix}} \sigma^2. \tag{51}$$

We now have an unbiased estimator, but unfortunately the uncertainty of $\hat{C}'_l$ has increased:

$$\mathrm{Var}^{1/2}(\hat{C}'_l) = \sqrt{\frac{2}{2l+1}} \left(C_l + \frac{4\pi}{N_{\rm pix}} \sigma^2 \right). \tag{52}$$

7.3. A DIGRESSION ON STATISTICAL METHODS IN GENERAL

The problem we just considered was a classic example of statistical parameter estimation. We had some *data*, $\{d\}$, from which we wanted to estimate a *parameter*, C_l. We did it by choosing an *estimator*, $\hat{C}'_l$, which we could compute from the data, and which we hoped would be close to the true value of the parameter.

In the problem above, there was a fairly natural choice of an estimator, but in general, for a more complicated problem, there may be no obvious choice. There is no universal, "correct" way to choose an estimator, but in many situations the *maximum-likelihood* estimator is a good choice. We will illustrate maximum-likelihood estimators with a simple example.

Suppose that we have M data points x_i, each of which is the sum of a signal s_i and some noise n_i. We will take both s_i and n_i to be Gaussian random variables with zero mean. The variances of the signal and noise are

$$\langle s_i^2 \rangle = S, \qquad\qquad \langle n_i^2 \rangle = N, \tag{53}$$

And everything is uncorrelated:

$$\langle s_i s_j \rangle = \langle n_i n_j \rangle = \langle s_i n_j \rangle = 0, \tag{54}$$

where the first two expressions assume $i \neq j$. Let us suppose we know the noise variance N, and we want to estimate the unknown quantity S, using a maximum-likelihood estimator.[11]

The first step is to compute the probability density of the data for fixed S. We want to know $p(\{x\} \mid S)$, where $p(\{x\} \mid S)\, d^M x$ is the probability of getting a set of data that lie within an infinitesimal volume $d^M x$ at the location of the actual data $\{x\}$. In this case, each x_i is an independent Gaussian with variance $S + N$,

$$p(x_i \mid S) = \frac{1}{\sqrt{2\pi(S+N)}} \exp\left(-x_i^2/2(S+N)\right), \tag{55}$$

and the joint probability density is the product

$$\begin{aligned} p(\{x\} \mid S) &= \prod_{i=1}^{M} p(x_i \mid s) && (56) \\ &= (2\pi(S+N))^{-M/2} \exp\left(\frac{-\sum_{i=1}^{M} x_i^2}{2(S+N)}\right). && (57) \end{aligned}$$

[11] The astute reader will have noticed that this is precisely the same problem we considered at the end of the last subsection. We have simply changed all of the notation for no good reason. To be specific, the correspondence with the previous problem goes like this: $M \to 2l+1$, $S \to C_l$, $s_i \to a_{lm}$, $N \to 4\pi\sigma^2/N_{\rm pix}$.

The probability density we have computed is a function of the data $\{x\}$ for fixed S. But the data are known, and S is what we want to know. We therefore choose to regard this probability density as a function of S and call it the *likelihood*.

$$L(S) = p(\{x\} \mid S). \tag{58}$$

When working with Gaussian probability distributions, it is often convenient to work with the quantity $\mathcal{L} \equiv -2 \ln L$ instead. The maximum-likelihood estimator, as its name suggests, is the value $\hat{S}$ of S for which L is maximized (or $\mathcal{L}$ is minimized). In other words, it is the value of the parameter for which it would have been most likely for us to get the data we actually did.

In the problem at hand, the maximum-likelihood estimator is found by differentiating

$$\mathcal{L} = M \ln 2\pi + M \ln(S + N) + \frac{\sum x_i^2}{S + N} \tag{59}$$

with respect to S, setting the result equal to zero, and solving for S. The result is

$$\hat{S} = \frac{1}{M} \sum_{i=1}^{M} x_i^2 - N. \tag{60}$$

That is, we compute the mean-square value of the data points and subtract of the noise bias. This is precisely what we did when we computed $\hat{C}_l'$ in equation (51). Although we didn't know it at the time, we were using a maximum-likelihood estimator.

In this case, the maximum-likelihood estimator turned out to be unbiased: its ensemble average $\langle \hat{S} \rangle$ is equal to the correct value S. In general, there is no guarantee that this will happen. To take a simple example, suppose that we had chosen to estimate the quantity S^{289} instead of S. The maximum-likelihood estimator would be $\hat{S}^{289}$, and it is easy to see that this quantity is highly positively biased.

Now we know how to estimate parameters. But in most cases an estimator isn't much good without a way of quantifying the uncertainty in it. Methods for doing this generally fall into two categories: the classical or *frequentist* approach (*e.g.*, Rice 1995) and the *Bayesian* approach (*e.g.*, Berger 1985, Gull & Daniell 1978, Press 1996). We will discuss each in turn.

In the frequentist picture, we look at one value of the parameter S at a time, and try to determine if that value is so far from our estimator $\hat{S}$ that it is ruled out. Specifically, for each S, we compute the probability distribution of the estimator $\hat{S}$. We use this probability distribution to determine how likely it is that we would have gotten a value of $\hat{S}$ as far off as we did, or worse. If the actual value of $\hat{S}$ is far off in the tail of the

probability distribution, then this probability will be low. If the probability lies below some *significance level* (say 5%), we say that that value of S is ruled out with 5% significance.[12] We repeat this process for a range of values of S, and we say that the set of values that are not ruled out form a 95% *confidence interval* for the parameter.

For a frequentist, a value of the parameter is ruled out if there is a low probability of getting data that fits as badly as the actual data. The Bayesian approach is quite different in spirit: A Bayesian attempts to determine the subjective probability distribution that characterizes her knowledge of the parameter given the data. Armed with that probability distribution, she can calculate how likely the parameter is to lie in any particular range.

In order to implement the Bayesian strategy, we want to turn the likelihood function $L(S) = p(\{x\} \mid S)$, which represents the probability of the data given a value of the parameter, into $p(S \mid \{x\})$, the probability of the parameter given the data. The way to do this is to invoke Bayes's theorem:

$$p(S \mid \{x\}) \propto p(\{x\} \mid S)p(S), \tag{61}$$

with the constant of proportionality chosen to make the integral of the left-hand side equal one. The left-hand side of this equation is the *posterior probability distribution*, and it is precisely what we are looking for: it tells us the probability of a particular parameter value, given the data. On the right-hand side we have the product of the likelihood function and the *prior distribution* of the parameter S. The latter represents our state of knowledge of S before we looked at the data.

A Bayesian characterizes the uncertainty in a parameter estimate by drawing a *credible region* around the estimate. A 95% credible region, for example, is an interval $S_{\min} < S < S_{\max}$ such that there is a 95% posterior probability that S lies in that interval,

$$\int_{S_{\min}}^{S_{\max}} p(S \mid \{x\})\, dS = 0.95. \tag{62}$$

The boundaries $S_{\min}$ and $S_{\max}$ of the credible region are typically chosen to have equal values of the posterior probability density.

Although the frequentist approach is the one most people think of when they think of statistics, and although most scientists profess to prefer it, many if not most error bars in cosmology are determined using Bayesian techniques.

[12] Astrophysicists often phrase that same statement differently, saying that the value is ruled out at *95% confidence*.

The main objection people raise to the Bayesian is that the final results depend on the prior distribution $p(S)$. For a true, orthodox Bayesian, this is not really a problem: the Bayesian view is that all probabilities represent our subjective knowledge, and that prior distributions are therefore secretly built into all statistical reasoning. It is better, the argument goes, to have the prior out in the open for all to see.

Whether or not you like this argument, there is no denying that in practice choosing a prior can be tricky. If one has essentially no prior knowledge about the parameter, then the prior distribution should be broad and flat. [For a flat prior, we can see from equation (61) that the posterior probability distribution is simply the likelihood function.] But even in this situation, it is not generally obvious which "flat" prior to choose. For example, if we are trying to estimate an element of the power spectrum C_l, should we choose a prior that is flat in C_l or one that is flat in $\sqrt{C_l}$? (C_l is after all a mean-square amplitude; maybe the r.m.s. amplitude is a more "natural" choice.) Perhaps we should even choose a prior that is flat in $\ln C_l$, since such a prior avoids choosing a preferred scale. It would be hard to say that any of these choices is "wrong," but in some situations the result of a calculation may depend on which choice is made. For an example, see Bunn *et al.* (1994).

The situation is not as bad as it appears, however. If the data set in question contain a good, strong detection of the parameter of interest, then the likelihood function is sharply peaked, and the shape of the posterior probability (61) is determined mostly by the likelihood rather than the prior. Prior dependence is thus typically weak in the case of strong detections. The situations where prior dependence is a serious problem are typically those in which someone is trying to coax a value out of a data set that is capable of only a weak constraint anyway.

7.4. INCOMPLETE SKY COVERAGE

We now return to our hypothetical CMB experiment. The next complication we need to consider has to do with the fact that no actual experiment ever achieves complete sky coverage. In the case of COBE, pixels close to the Galactic plane are contaminated, leaving only about two thirds of the sky usable. All other experiments to date have covered even smaller patches of sky.

This fact requires us to completely change our approach. As much as we would like to estimate each a_{lm} and hence each C_l individually, in the absence of complete sky coverage it is impossible to do so. There is in fact no estimator of a particular C_l that is "uncontaminated," *i.e.*, that is independent of all of the other $C_{l'}$.

We may decide that it is important to estimate each C_l individually, with the minimum possible contamination from other multipoles. Tegmark (1996a) has devised power-spectrum estimators with this property in mind and has applied them to both galaxy surveys (Tegmark 1995) and the four-year COBE data (1996b). For instance, suppose we have our hearts set on knowing the value of C_{17} as well as possible. Since the power spectrum is quadratic in $\Delta T/T$, it is natural to choose a quadratic estimator,

$$\hat{C}_{17} = \sum_{i,j} A_{ij} d_i d_j - B. \tag{63}$$

Here d_i is a data point and we want to choose the matrix elements A_{ij} and the bias correction B in order to get as good an estimator as possible. Tegmark (1996a) proposes that we choose these quantities to make our estimator unbiased and to minimize the dependence of $\hat{C}_{17}$ on all of the other C_l's. He shows that it is impossible to completely remove contamination from other multipoles and that in general the "spectral resolution" Δl of an experiment is approximately the reciprocal of the angular scale $\Delta\theta$ covered by the sky map. In particular, for an experiment like COBE, $\Delta\theta \sim 1$ radian, and it turns out that it is possible to estimate a particular C_l with significant contamination only from modes with $\Delta l \approx 2$ (Tegmark 1996a, 1996c).

7.5. MAXIMUM-LIKELIHOOD PARAMETER ESTIMATION

We may, however, decide that it isn't so important to estimate each C_l individually. Often, a more fruitful approach is to parameterize the power spectrum C_l with a small number k of parameters,

$$C_l = C_l(q_1, q_2, \ldots, q_k), \tag{64}$$

and use maximum-likelihood methods to estimate those parameters. This is in fact the usual approach in CMB data analysis. Specific choices of the parameters $\{q\}$ include the following:

- We may assume a *shape* for the power spectrum and estimate the *normalization*. In this case, there is only one free parameter, which is conventionally taken to be the quadrupole amplitude $\langle Q \rangle \equiv \sqrt{5C_2/4\pi}$.[13] Most degree-scale experiments are only powerful enough to determine a single number, the total power. One therefore frequently assumes a

[13] A bewildering variety of notations exist in the literature. We choose to call this quantity $\langle Q \rangle$ to emphasize the fact that it is a theoretical ensemble-average quantity. In particular, it is not the same as the local quadrupole $Q_{\rm rms} \equiv \sum_{m=-l}^{l} |a_{2m}|^2/4\pi$. The COBE group generally denotes its estimators of $\langle Q \rangle$ by $Q_{\rm rms-PS}$.

"flat" power spectrum $l(l+1)C_l = \text{const.}$ and estimates the normalization, which in this context is often called $Q_{\rm flat}$.

- Both the normalization $\langle Q \rangle$ and the spectral index n may be chosen as free parameters. For a large-angle experiment like COBE, the predicted power spectrum depends only weakly on many of the other parameters.
- White & Bunn (1995) have suggested a phenomenological parameterization of the power spectrum. At large angular scales, many popular theoretical models are well approximated by power spectra that are quadratics in $\log l$. To be specific, we may set

$$l(l+1)C_l = D_1(1 + D'(\log_{10} l - 1) + \tfrac{1}{2}D''(\log_{10} l - 1)^2) \tag{65}$$

  and work with a three-parameter family (D_1, D', D'') of power spectra.
- We may choose to divide the power spectrum over the range probed by a particular experiment into a small number of "bands." We then estimate the power in each band, assuming that $l(l+1)C_l$ is constant in each band. This has been done for COBE (Hinshaw *et al.* 1996) and Saskatoon (Netterfield *et al.* 1996), although the latter uses a completely different method.

No matter what parameterization we adopt, we need a way to compute the likelihood L for a given power spectrum. As long as we assume Gaussian statistics, it is relatively easy to write down a formula for the likelihood, although as we shall see it can be cumbersome to compute it in practice.

We begin by introducing some notation. Each data point d_i is as usual the sum of the signal $\Delta T/T(\hat{\mathbf{r}}_i)$ and noise n_i. Expanding $\Delta T/T$ in spherical harmonics, we have

$$d_i = \sum_{l,m} a_{lm} Y_{lm}(\hat{\mathbf{r}}_i) + n_i. \tag{66}$$

Let us denote a pair of indices (lm) by a single Greek index μ. The correspondence is $\mu = l(l+1) + m$, so that μ ranges from 1 to ∞ as (lm) take on all of their allowed values. Then we can write equation (66) more compactly as

$$\vec{d} = \mathbf{Y}\vec{a} + \vec{n}, \tag{67}$$

where $\vec{d} = (d_1, d_2, \ldots, d_{N_{\rm pix}})$ is the data vector,[14] $\vec{n} = (n_1, \ldots, n_{N_{\rm pix}})$ is the noise vector, and the infinite-dimensional vector $\vec{a} = (a_1, a_2, \ldots, a_\mu, \ldots)$ contains the spherical harmonic coefficients. The $N_{\rm pix} \times \infty$-dimensional spherical harmonic matrix $\mathbf{Y}$ has elements

$$Y_{i\mu} = Y_\mu(\hat{\mathbf{r}}_i). \tag{68}$$

[14]We denote vectors that live in abstract spaces such as "pixel space" by arrows, and vectors in real three-dimensional space are written in boldface.

The statistical properties of $\vec{d}$ are determined by the properties of $\vec{a}$ and $\vec{n}$. Assuming Gaussian statistics, both are Gaussian random vectors with zero mean and covariances given by

$$\langle a_\mu a_\nu^* \rangle = C_\mu \delta_{\mu\nu} \equiv C_{\mu\nu}, \tag{69}$$

$$\langle n_i n_j \rangle = \sigma_j^2 \delta_{ij} \equiv N_{ij}, \tag{70}$$

$$\langle a_\mu n_i \rangle = 0. \tag{71}$$

($C_\mu \equiv C_l$ where l is the index corresponding to μ, and $\mathbf{C}$ and $\mathbf{N}$ are diagonal matrices.) Since $\vec{d}$ is a linear combination of $\vec{a}$ and $\vec{n}$, it too is a multivariate Gaussian, and the likelihood function therefore has the form

$$L(C_l) \equiv p(\vec{d} \mid C_l) = \frac{1}{(2\pi)^{N_{\rm pix}/2} \det^{1/2} \mathbf{M}} \exp\left(-\tfrac{1}{2}\, \vec{d}^{\rm T} \mathbf{M}^{-1} \vec{d}\right). \tag{72}$$

The T denotes a transpose, and the covariance matrix $\mathbf{M}$ is given by

$$\mathbf{M} \equiv \langle \vec{d}\vec{d}^{\rm T} \rangle = \langle (\mathbf{Y}\vec{a} + \vec{n})(\mathbf{Y}\vec{a} + \vec{n})^{\rm T} \rangle = \mathbf{Y}\mathbf{C}\mathbf{Y}^{\rm T} + \mathbf{N}. \tag{73}$$

In principle, we are now ready to estimate parameters. Equation (72) tells us how to compute the likelihood for any particular power spectrum C_l, so all we need to do is hunt through our parameter space for the parameters that maximize the likelihood.

In fact, for a typical degree-scale experiment with tens or at most hundreds of pixels, this is essentially what is done. For a large data set such as COBE, though, there are too many pixels for this to be convenient: each time we wish to compute a likelihood, we must invert the $N_{\rm pix} \times N_{\rm pix}$ matrix $\mathbf{M}$. For COBE, therefore, we must implement some form of "data compression" to make the analysis tractable. (Data compression will be even more essential for a future satellite experiment with orders of magnitude more pixels than COBE.)

7.6. BEAM-SMOOTHING AND CHOPPING

Before we discuss data compression, though, we need to discuss one more issue. The hypothetical experiment we have been discussing is still overly idealized in one important way. We have assumed that the signal measured by the experiment is the temperature anisotropy $\Delta T/T$ at a point. In reality, no experiment has perfect resolution, so the observed signal is actually the convolution of $\Delta T/T$ with some beam pattern or point-spread function. Furthermore, many experiments chop their beams between two (or more) points on the sky, with the measured signal being a difference between these points.

The effect of the beam pattern on our analysis is fairly simple. Let $B(\alpha)$ represent the response of the instrument to a point an angular distance α from the line of sight. (We assume that the beam pattern is azimuthally symmetric.) Then what the experiment actually measures is the convolution of the anisotropy with the beam pattern,

$$\left(\frac{\Delta T}{T} \star B\right) = \sum_{l,m} \bar{a}_{lm} Y_{lm}. \tag{74}$$

The coefficients $\bar{a}_{lm}$ are related to the true anisotropy coefficients a_{lm} like this:[15]

$$\bar{a}_{lm} = B_l a_{lm}, \tag{75}$$

where B_l is the expansion in Legendre polynomials of B,

$$B_l = \int_{-1}^{1} d\cos\alpha \, B(\alpha) P_l(\cos\alpha). \tag{76}$$

If the beam pattern happens to be a Gaussian,

$$B(\alpha) \propto \exp(-\alpha^2/2\sigma^2), \tag{77}$$

then the Legendre coefficients are

$$B_l = \exp(-\tfrac{1}{2}\sigma^2 l(l+1)). \tag{78}$$

Note that as expected B_l is very small for $l \gg \sigma^{-1}$, *i.e.*, for angular scales $\theta \ll \sigma$.

We can adapt all of the previous results of this section to take beam-smoothing into account by simply saying that our experiment is measuring the beam-smoothed power spectrum,

$$\bar{C}_l \equiv C_l B_l^2, \tag{79}$$

instead of C_l.

We can account for the effect of beam-switching in a similar way. Consider an experiment that chops between two points with spherical coordinates $(\theta, \phi + \frac{1}{2}\alpha)$ and $(\theta, \phi - \frac{1}{2}\alpha)$. Ignoring beam-smoothing, the observed signal d is the difference in the anisotropy between these two points:

$$d = \frac{\Delta T}{T}(\theta, \phi + \tfrac{1}{2}\alpha) - \frac{\Delta T}{T}(\theta, \phi - \tfrac{1}{2}\alpha) \tag{80}$$

$$= \sum_{l,m} a_{lm} \left(Y_{lm}(\theta, \phi + \tfrac{1}{2}\alpha) - Y_{lm}(\theta, \phi - \tfrac{1}{2}\alpha) \right). \tag{81}$$

[15]This result is simply the spherical version of the convolution theorem for Fourier transforms, $\widetilde{f \star g} = \tilde{f}\tilde{g}$.

The azimuthal dependence of Y_{lm} is $\exp(im\phi)$, so

$$d = \sum_{l,m} a_{lm} Y_{lm}(\theta,\phi) \left(\exp(\tfrac{1}{2} im\alpha) - \exp(-\tfrac{1}{2} im\alpha) \right) \quad (82)$$

$$= \sum_{l,m} a_{lm} Y_{lm}(\theta,\phi) 2i \sin \tfrac{1}{2} m\alpha. \quad (83)$$

The net result is that a_{lm} is replaced by $2ia_{lm} \sin \frac{1}{2} m\alpha$, so modes with low $|m|$ are suppressed. Since m ranges from $-l$ to l, this suppression affects primarily modes with low l.[16]

This suppression is conventionally quantified by computing a "window function" that represents the sensitivity of the experiment to different multipoles. To do this, we compute the mean-square signal,

$$\langle d^2 \rangle = \sum_{l,m} C_l |Y_{lm}(\theta,\phi)|^2 \left(2 \sin \tfrac{1}{2} m\alpha \right)^2 \equiv \sum_l \left(\frac{2l+1}{4\pi} \right) C_l W_l. \quad (84)$$

The window function W_l is small for low l, indicating that chopping has rendered this experiment insensitive to the largest angular scales.

Note that we have not included beam-smoothing in equation (84). The correct window function, including beam-smoothing, is obtained by multiplying this result by B_l^2.

Equation (84) gives the window function for the particularly simple case of a single-difference experiment. There are more complicated switching strategies, including sinusoidal chops and triple-beam experiments. For a more detailed discussion of window functions, see White & Srednicki (1995).

8. Likelihood Analysis of the COBE Data

In the previous section, we discussed various issues of CMB data analysis from a general point of view. We will now apply what we have learned to a specific example, namely the COBE DMR data. We will not describe the COBE instrument in detail; the interested reader is referred to George Smoot's contribution to this volume, as well as to the papers reporting the four-year DMR data (Bennett *et al.* 1996, Górski *et al.* 1996, Hinshaw *et al.* 1996, Banday *et al.* 1996) and references therein. We will content ourselves with mentioning a few of the most relevant facts.

The COBE DMR produced all-sky maps of the microwave radiation at three frequencies, 31 GHz, 53 GHz, and 90 GHz, with a beam size of 7°

[16]The fact that modes with low $|m|$ are suppressed depends on the fact that we have oriented our coordinate system with the chop in the azimuthal direction. In contrast, the statement that, on average, modes with low l are suppressed is independent of the orientation of the coordinate system.

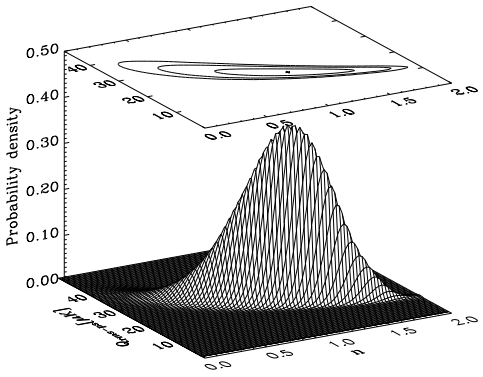

Figure 11. The likelihood function for the two-year COBE DMR data, based on a brute-force analysis involving the entire pixel covariance matrix. Only the Sachs-Wolfe contribution to the anisotropy is included. See Tegmark & Bunn (1995) for further details.

(FWHM). The maps consist of 6144 pixels, although only about 4000 of them are at high enough Galactic latitude to be used for studying the CMB. Although the DMR is a differencing instrument, the data have been used to produce sky maps of $\Delta T/T$, so we do not need to worry about beam-switching in our analysis. We do, however, have to worry about the fact that the maps are insensitive to the monopole and dipole of the anisotropy.[17]

The noise in the COBE maps appears to be Gaussian, and different pixels have noise that is approximately uncorrelated (Lineweaver *et al.* 1994). Therefore, as long as the CMB anisotropy obeys Gaussian statistics, equation (72) applies:

$$\mathcal{L} \equiv -2\ln L = \ln\left((2\pi)^{N_{\rm pix}}\det\mathbf{M}\right) + \vec{d}^{\rm T}\mathbf{M}^{-1}\vec{d}, \tag{85}$$

where $\mathbf{M} = \mathbf{Y}\bar{\mathbf{C}}\mathbf{Y}^{\rm T} + \mathbf{N}$ with $N_{ij} = \sigma_i^2\delta_{ij}$. The matrix $\mathbf{M}$ is $\sim 4000 \times 4000$, which is a size that can be inverted, with sufficient patience, on a workstation. Tegmark & Bunn (1995) have performed such a brute-force analysis on the two-year COBE DMR data for a two-parameter family of power spectra, with results shown in Figure 11. However, if we wish to explore a larger parameter space, we must find a more efficient way to compute likelihoods.

[17] Actually, COBE is in principle perfectly sensitive to the dipole; however, the intrinsic CMB dipole is impossible to distinguish from the much larger dipole due to our own motion with respect to the CMB center-of-momentum frame.

8.1. DATA COMPRESSION

All likelihood analyses of the COBE data, with the exception of the brute-force analysis mentioned above, have involved some form of data compression. That is, the pixel data $\vec{d}$ has been mapped to some smaller-dimensional data vector, which has been used for computing likelihoods. We will focus on *linear* methods of data compression, in which the compressed data vector $\vec{x}$ is linear in $\vec{d}$,

$$\vec{x} = \mathbf{A}\vec{d}, \tag{86}$$

for some $K \times N_{\rm pix}$ matrix $\mathbf{A}$ with $K < N_{\rm pix}$. $\vec{x}$ is a Gaussian random vector, so we can use equation (72) to compute the likelihood of $\vec{x}$ in terms of the covariance matrix

$$\mathcal{M} \equiv \langle \vec{x}\vec{x}^{\rm T} \rangle = \mathbf{A}\mathbf{M}\mathbf{A}^{\rm T}. \tag{87}$$

Of course, the likelihood computed in this way will not be the same as the true likelihood computed from $\vec{d}$, but we can hope that, if we perform our data compression wisely, we will get a reasonable approximation to the true likelihood.

In effect, linear data compression is equivalent to expanding the sky map in a set of normal modes, namely the rows of $\mathbf{A}$. Each element of the compressed data vector $\mathbf{x}$ is approximately the integral of the sky map, multiplied by some function: heuristically, we can write

$$x_i = \sum_{j=1}^{N_{\rm pix}} A_{ij} d_j \approx \int A_i(\hat{\mathbf{r}}) d(\hat{\mathbf{r}})\, d\Omega. \tag{88}$$

If our pixels uniformly covered the whole sky, we would choose these mode functions to be the spherical harmonics by setting $A_{\mu j} = Y_\mu(\hat{\mathbf{r}}_j)$. Then x_μ would be an estimate of a_μ (up to an overall normalization). In fact, we would be performing precisely the analysis described in Section 7.2.

Even though we do not actually have complete sky coverage, there is still nothing stopping us from choosing the rows of $\mathbf{A}$ to be the spherical harmonics. This is in fact the technique described by Górski (1994), which has been applied to the DMR data by Górski *et al.* (1994, 1996).[18] By cutting off the spherical harmonic expansion at $l = 30$, Górski *et al.* compress the data from ~ 4000 to ~ 1000 numbers, with little loss of cosmological information. This is possible because the cosmic signal in the data drops off rapidly with increasing l (due to both the beam cutoff and the shape of

[18] Górski's method involves the additional step of orthogonalizing the spherical harmonics with an algorithm like Gram-Schmidt. Orthogonalizing with respect to the monopole and dipole is an excellent way to render the data insensitive to these modes, but orthogonalizing the modes with $l \geq 2$ with respect to each other has no effect on the likelihoods.

the anisotropy power spectrum), while the noise has approximately equal power in all modes.

8.2. THE KARHUNEN-LOÈVE TRANSFORM

The Karhunen-Loève transform (Karhunen 1947), which is also known as optimal subspace filtering or expansion in signal-to-noise eigenmodes, is another prescription for linear data compression. It was first introduced to CMB data analysis by Bond (1994, 1995, 1996), and has been used extensively on the COBE data (Bunn, Scott, & White 1995; Bunn 1995; White & Bunn 1995; Bunn & Sugiyama 1996; Bunn, Liddle, & White 1996; Bunn & White 1996) as well as in analyzing galaxy catalogues (Vogeley & Szalay 1996).

Let us consider a one-parameter family of power spectra $C_l(q)$, where the true value of q is q_0. We wish to choose our method of data compression (*i.e.*, the matrix $\mathbf{A}$) to enable us to estimate q as well as possible. Specifically, we choose $\mathbf{A}$ to maximize our ability to reject incorrect values of q.

On average, the likelihood function $L(q)$ has a peak at the true value $q = q_0$, so $\langle L'(q_0)\rangle = 0$. The average rejection power is determined by the rate at which the likelihood declines when we move away from this peak. The figure of merit for describing rejection power is therefore

$$\gamma \equiv \left\langle \left. \frac{d^2\mathcal{L}}{dq^2} \right|_{q=q_0} \right\rangle . \tag{89}$$

The Karhunen-Loève transform consists of choosing the compression matrix $\mathbf{A}$ to maximize γ (for a fixed value of K, the dimension of the compressed data vector).

To solve this optimization problem, we write down the likelihood in terms of the reduced data vector $\vec{x}$,

$$\mathcal{L} = K \ln 2\pi + \mathrm{Tr}\left(\ln(\mathbf{AMA}^{\mathrm{T}}) + (\mathbf{AMA}^{\mathrm{T}})^{-1}\vec{x}\vec{x}^{\mathrm{T}} \right). \tag{90}$$

Then we compute γ, vary a matrix element A_{ij}, and set $\delta\gamma = 0$. After some algebra, we find that each row $\vec{\alpha}_a$ of $\mathbf{A}$ must satisfy an eigenvalue equation,

$$\mathbf{M}_0'\vec{\alpha}_a = \lambda_a \mathbf{M}_0 \vec{\alpha}_a. \tag{91}$$

Here $\mathbf{M}_0$ is the covariance matrix $\mathbf{M}$ corresponding to the correct parameter value $q = q_0$, and

$$\mathbf{M}_0' = \left. \frac{d\mathbf{M}}{dq} \right|_{q=q_0} . \tag{92}$$

The rejection power γ is simply the sum of the squares of the eigenvalues λ_a.

This completes our prescription for choosing the matrix $\mathbf{A}$. We should choose the rows of $\mathbf{A}$ to be the solutions of equation (91) with the largest values of $|\lambda_a|$. Furthermore, we know when it is safe to stop adding new rows: once all of the remaining eigenvalues λ_a are small, we will no longer significantly increase γ by adding more rows to $\mathbf{A}$.

To get an intuitive understanding of the Karhunen-Loève transform, consider the case where the parameter q is the normalization of the power spectrum, so $C_l(q) = qC_l^{(0)}$. Then we can rewrite the eigenvalue equation (91) as

$$\mathbf{M}_{\rm signal}\vec{\alpha}_a = \hat{\lambda}_a \mathbf{M}_{\rm noise}\vec{\alpha}_a, \tag{93}$$

where $\hat{\lambda}_a = \lambda_a/(1-\lambda_a)$, and $\mathbf{M}_{\rm signal} = \mathbf{Y}\bar{\mathbf{C}}\mathbf{Y}^{\rm T}$ and $\mathbf{M}_{\rm noise} = \mathbf{N}$ are the signal and noise contributions to $\mathbf{M}$. We can see from equation (93) that $\vec{\alpha}_a$ is an eigenvector of $\mathbf{M}_{\rm noise}^{-1}\mathbf{M}_{\rm signal}$. This is why Bond (1994, 1995) calls it an "eigenmode of the signal-to-noise ratio." In effect, the Karhunen-Loève transform tells us which directions in the $N_{\rm pix}$-dimensional pixel space are most sensitive to the cosmic signal, and which are dominated by noise.

The reader, being an extraordinarily perceptive soul, is no doubt wondering at this point whether this whole procedure is worth the trouble. After all, our original goal was to avoid having to invert an $N_{\rm pix} \times N_{\rm pix}$ matrix. Now we find ourselves having to solve an $N_{\rm pix}$-dimensional eigenvalue problem, which is much harder than simply inverting a matrix. Recall, however, that our objection to a brute-force likelihood analysis was that we didn't want to invert the large matrix $\mathbf{M}$ *repeatedly* as we varied the power spectrum. The Karhunen-Loève eigenvalue problem needs to be solved only once, with all future operations being performed on the K-dimensional compressed data vector. Furthermore, it turns out that we can save ourselves a lot of work by solving equation (91) in spherical harmonic space rather than real space (Bunn 1995). Once we choose some cutoff $l_{\rm max}$, the dimension of the eigenvalue problem is reduced from $N_{\rm pix}$ to $\sim l_{\rm max}^2$. It turns out that none of the high signal-to-noise eigenmodes have significant power beyond $l = 30$ or so, so we can safely choose $l_{\rm max}$ to be 40 or 50, resulting in a substantial saving in computational effort.

The Karhunen-Loève transform depends on a choice of power spectrum. Ideally, we would like to use the true power spectrum, but of course we don't know the true power spectrum.[19] We must therefore choose a *fiducial power spectrum* more or less arbitrarily. In principle, this could lead to trouble: we might find that the choice of fiducial power spectrum had a significant effect on our final results. There are two ways to address this question: we can repeat the analysis with different fiducial power spectra, and we

[19] If we did, there would be no need to perform the analysis!

can perform Monte Carlo simulations to check that the likelihood analysis returns unbiased estimates of the parameters of interest.[20]

In the case of the COBE data, extensive tests have revealed that sensitivity to the fiducial power spectrum is not a problem (Bunn 1995, Bunn & White 1996). For example, the maximum-likelihood normalization of an $n = 1$ Sachs-Wolfe spectrum is $\langle Q \rangle = 18.73 \pm 1.25\,\mu$K using an $n = 1$ fiducial power spectrum and $\langle Q \rangle = 18.74 \pm 1.25\,\mu$K using an $n = 1.5$ power spectrum. The maximum-likelihood value of n also does not change when we change the fiducial power spectrum. Furthermore, Monte Carlo simulations show that our estimates of $\langle Q \rangle$ and n are unbiased to an accuracy much better than the statistical uncertainty ($\lesssim 0.03\,\mu$K and $\lesssim 0.05$ respectively). See Bunn & White (1996) for further details.

8.3. MONOPOLE AND DIPOLE REMOVAL

Since the COBE data do not contain useful monopole and dipole information, it is customary to remove a best-fit monopole and dipole from the data before performing any further analysis. Unfortunately, since incomplete sky coverage destroys the orthogonality of the spherical harmonics, this procedure covertly removes part of the contribution of the higher multipoles. There are two ways to compensate for this.

The first option is to treat the monopole and dipole coefficients (a_{00} and a_{1m}) as "nuisance parameters," *i.e.*, quantities whose true values we neither know nor care about.[21] In the context of Bayesian analysis, the natural thing to do with nuisance parameters it to marginalize over them. Marginalizing over a nuisance parameter ζ means replacing the likelihood L with the marginal likelihood

$$L_{\rm marg} = \int d\zeta L(\zeta) p(\zeta). \tag{94}$$

Here $p(\zeta)$ is a prior probability density for ζ, which is usually taken to be constant. By marginalizing over the data, we are using a standard identity of probability theory,

$$p(x) = \int p(x \mid \zeta) p(\zeta)\, d\zeta, \tag{95}$$

[20]Even in methods that do not involve a choice of fiducial power spectrum, it is wise to perform simulations to test for bias. Even a brute-force likelihood analysis using the full pixel data is not guaranteed to return unbiased parameter estimates.

[21]A parameter may be a nuisance parameter at one time and an interesting parameter at another. For instance, if we want to estimate the spectral index n, we should probably compute $L(\langle Q \rangle, n)$ and treat $\langle Q \rangle$ as a nuisance parameter. At some other time, though, we may think $\langle Q \rangle$ is an interesting thing to know.

to remove all ζ-dependence from the likelihood.

From a frequentist point of view, the natural way to get rid of a nuisance parameter is to maximize with respect to it. That is, we replace L with $\max_\zeta L(\zeta)$. That way, a particular model is ruled out only if it is ruled out for all possible values of ζ.

If we are performing some sort of data compression, then we have a second option for dealing with the monopole and dipole. We can simply impose a constraint on our compression matrix $\mathbf{A}$, requiring it to be insensitive to the unwanted multipoles. This is in effect the approach of Górski (1994): by orthogonalizing the spherical harmonics, he makes his compression matrix insensitive to the monopole and dipole. This approach turns out to be mathematically equivalent to marginalizing over the unwanted modes.

People frequently remove the quadrupole information from the COBE data in the same way as the monopole and dipole, on the grounds that the quadrupole is particularly susceptible to Galactic contamination. It has also been known since the earliest days of COBE analysis that the quadrupole is anomalously low (compared to the prediction of a flat power spectrum normalized to the other multipoles). From a statistical point of view, this is a delicate situation: it is perfectly acceptable, and even wise, to throw away data if there is a reasonable fear of contamination, but throwing away data that is known *a priori* to be discordant with favored theories is a major statistical *faux pas*. On balance, it is probably better to leave the quadrupole information in in the interest of avoiding even the possibility of biased editing of the data.

There is another argument in favor of retaining the quadrupole. Even if the quadrupole is contaminated, it still contains useful information, and so it may be unwise to throw it away entirely. Since the quadrupole is a root-mean-square quantity, any contaminant would tend to bias the quadrupole up. In fact, if a particular theory is ruled out because it predicts too large a quadrupole, hypothesizing an additional quadrupolar contaminant cannot save that theory: as long as the contaminant is statistically independent of the cosmic signal, the net result of hypothesizing a contaminant is necessarily to *lower* the likelihood of that theory.

8.4. RESULTS

The main purpose of this section is to discuss data analysis techniques, not results; however, we will briefly present some results based on a Karhunen-Loève analysis of the four-year COBE DMR data. The reader is referred to Bunn & White (1996) for a more detailed discussion.

The data set used for this analysis consists of a weighted average of the

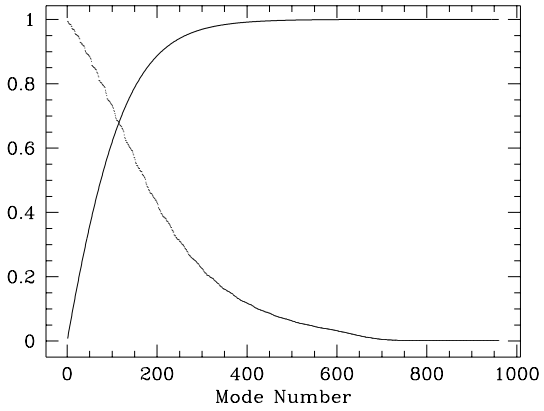

Figure 12. The points show the eigenvalues λ_a of the four-year COBE data, sorted in decreasing order. The solid curve is the running sum of λ_a^2, normalized to 1.

53 and 90 GHz maps from the four-year DMR data. The maps are averaged with weights inversely proportional to the noise variance, in order to minimize noise in the average map. (This is equivalent to performing a joint likelihood analysis of the individual maps.) We performed the Karhunen-Loève analysis using a flat fiducial power spectrum $l(l+1)C_l = \text{const.}$, and we retained the 500 most significant modes.

Figure 12 shows the eigenvalues λ_a, together with a running sum of the squares of the eigenvalues. (Recall that this sum is proportional to the rejection power γ.) This plot indicates that modes beyond the first 500 do not significantly increase our ability to discriminate among models.

Figure 13 shows the likelihood function for low-density CDM models, both with and without a cosmological constant. Figure 14 shows the maximum-likelihood power spectrum, found by allowing each C_l with $2 \leq l \leq 19$ to vary independently. The error bars shown in this figure are standard errors determined by approximating the likelihood near the peak as a Gaussian. The standard errors are then the square roots of the diagonal elements of the covariance matrix of this Gaussian. Error bars determined in this way should be viewed with extreme caution. First, the likelihood is not very well approximated by a Gaussian: on the contrary, it is strongly skew-positive at low l. Second, these standard errors contain no information about correlations between the errors. These correlations are largest for pairs of modes whose l-values differ by 2. (Coupling between modes with $\Delta l = 1$ is weak because the data have approximate reflection symmetry.) The deceptively small error bar on the estimate of C_2 is largely due to the failure of the Gaussian approximation for the likelihood, although the 15%

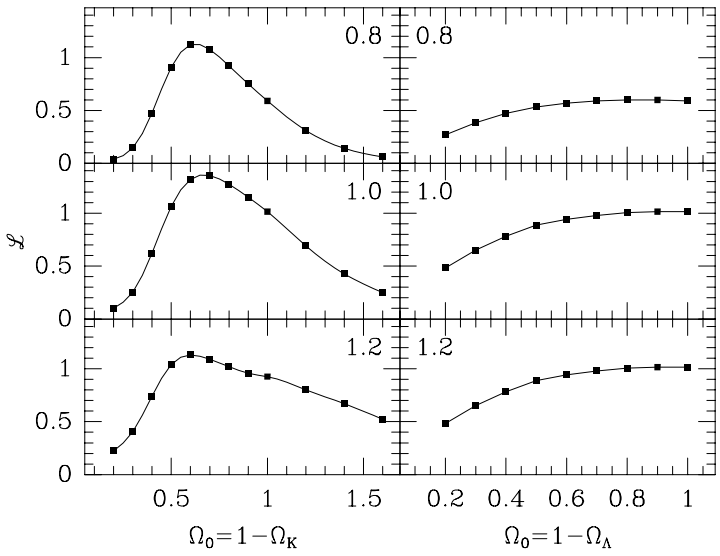

Figure 13. Likelihood as a function of Ω_0 for CDM models with zero cosmological constant (left) and zero spatial curvature (right). The spectral index n increases from 0.8 to 1.2 from top to bottom. The likelihoods are normalized so that a flat spectrum has $L = 1$. See Bunn & White (1996) for further details.

anticorrelation between C_2 and C_4 also plays a role.

Finally, Table 1 shows values of the small-scale fluctuation amplitude σ_8 for various theoretical models. The observational constraint is approximately $0.5 \lesssim \sigma_8 \lesssim 0.8$ (*e.g.*, Viana & Liddle 1996).

8.5. WIENER FILTERING

Until now, we have focused on attempts to estimate the angular power spectrum C_l. While this is the most useful thing to do with a CMB data set, other complementary approaches can be interesting in certain contexts. For instance, we can assume that we know the angular power spectrum and try to determine the underlying cosmic signal from a noisy sky map. That is, we can attempt to *filter* a sky map, cleaning up the noise and leaving the signal. The Wiener filter (Wiener 1949) is optimal linear filter for this purpose, in the sense of least squares. The recent use of Wiener filtering in astrophysics is largely due to Rybicki & Press (1992), and the filter has been applied to the COBE data by Bunn, Hoffman, & Silk (1996).

Suppose we have a data vector $\vec{d}$ containing signal and noise. We want to apply a linear filter $\mathbf{F}$ so that $\vec{y} \equiv \mathbf{F}\vec{d}$ approximates the true cosmic

	Ω_0	Ω_Λ	$\Omega_{\rm HDM}$	n	h	$\Omega_{\rm B}h^2$	σ_8
standard CDM	1.0	0.0	0.0	1.0	0.50	0.0125	1.22
tilted CDM	1.0	0.0	0.0	0.8	0.50	0.0250	0.72
MDM	1.0	0.0	0.2	1.0	0.50	0.0150	0.79
ΛCDM	0.4	0.6	0.0	1.0	0.65	0.0150	1.07
Open CDM	0.4	0.0	0.0	1.0	0.65	0.0150	0.64
Low h CDM	1.0	0.0	0.0	1.0	0.35	0.0150	0.74

TABLE 1. The predicted fluctuation amplitude on scales of $8\,h^{-1}$Mpc for various CDM-like models. MDM is a "mixed dark matter" model. All normalizations are from the four-year COBE DMR data. See Bunn & White (1996) for further details.

signal $\Delta T/T$ in such a way that the mean-square deviation,

$$\left\langle \left(y_i - \frac{\Delta T}{T}(\hat{\mathbf{r}}_i) \right)^2 \right\rangle, \tag{96}$$

is as small as possible. The solution to this optimization problem is the Wiener filter,

$$\mathbf{F} = \mathbf{M}_{\rm signal}\mathbf{M}^{-1}, \tag{97}$$

where $\mathbf{M}$ is as usual the data covariance matrix and $\mathbf{M}_{\rm signal}$ is the signal contribution to $\mathbf{M}$.

Under the assumption of Gaussian statistics, the Wiener-filtered data is also the maximum-likelihood estimator of $\Delta T/T$ at each point. Note that in regions of very high noise, where we have little information, the Wiener filter returns values near zero, because this is the most likely *a priori* value of a zero-mean Gaussian.

Figure 15 shows a Wiener-filtered COBE sky map. Although the signal-to-noise ratio in the raw pixel maps is typically less than one per pixel, the largest-amplitude features in the filtered map are significant at the five sigma level per pixel.

One of the main uses of the filtered maps is in making predictions for other experiments. Assuming Gaussian statistics, the full error covariance matrix of the Wiener-filtered map is known, and so we can produce maps with known uncertainties of a region of the sky. For predictions of the CMB sky as it should be seen by the Tenerife experiment, see Bunn, Hoffman, & Silk (1996).

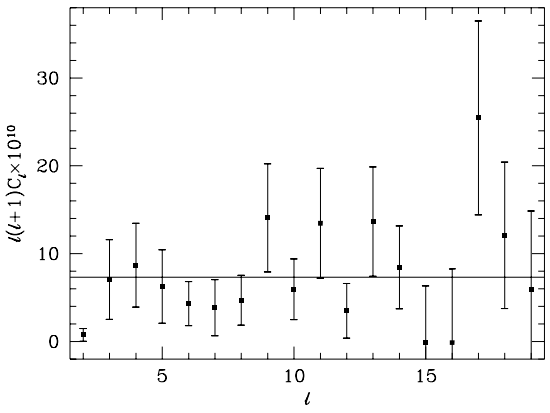

Figure 14. The points represent the maximum-likelihood power spectrum, obtained by letting all C_l's between 2 and 19 vary freely. A flat $\langle Q \rangle = 19\,\mu$K power spectrum is plotted for comparison. The error bars are standard errors determined by approximating the likelihood by a Gaussian near the peak. Because the Gaussian approximation is poor, and because there are significant correlations between the errors, these error bars can be deceptive. The small formal error on C_2 is particularly misleading. See Bunn & White (1996) for further discussion.

9. Summary

The main lesson to be learned from this entire institute is that this is an exciting time in CMB research. The existing data are already telling us vast amounts about cosmology, and in the next few years the data should continue to improve dramatically. The high quality of present and future anisotropy observations presents us with some challenges. We must understand our theoretical models well enough to make accurate predictions, and we must develop statistical tools that enable us to determine which predictions are consistent with the data. Both of these challenges are currently being met with ever-increasing success.

The tools for making accurate predictions, at least in linear models like CDM, are by now quite well developed. Furthermore, in recent years analytic and semianalytic approximations have dramatically improved our understanding of the basic physical principles involved in anisotropy formation.

The problem of data analysis is also much better understood today than it was five years ago (before there were any actual detections to analyze). However, it is important to remember that analysis of future data sets will present challenges that make the COBE analysis look easy. When sky maps contain a million pixels instead of a few thousand, data compression will be absolutely essential. It is already time to start thinking about this difficult

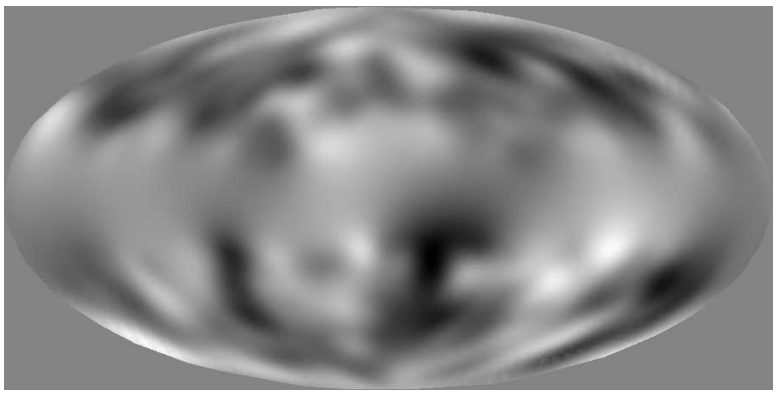

Figure 15. A sky map, in Aitoff projection, of the Wiener-filtered four-year DMR data. The relative lack of structure near the Galactic plane is due to the fact that no data from that region were used.

problem.

In addition, future experiments with higher resolution than COBE will be more susceptible to foreground contamination. In the case of COBE, it is believed that simply excising points too close to the Galactic plane is sufficient to remove most of the foreground contamination; for future high-sensitivity degree-scale experiments, more sophisticated methods will be necessary.

We have seen that CDM-like theoretical models predict that vast amounts of information are encoded in the CMB anisotropy power spectrum. There is a very real hope that the CMB will give us accurate values for all sorts of cosmological parameters. But even if the information is there, we will have to do a lot of work to wrest it from the data.

10. Acknowledgments

The first half of these lectures is based largely on the work of Wayne Hu and Naoshi Sugiyama. Much of the later material is based on work I performed in collaboration with Douglas Scott, Joseph Silk, Max Tegmark, and Martin White. I would like to thank all of these people for many helpful discussions. In addition, Wayne, Max, and Martin made some of the figures. Finally, I would like to thank the organizers of the meeting for their hard work and hospitality.

References

1. Banday, A. *et al.* 1996, preprint (astro-ph/9601065).
2. Bennett, C.L. *et al.* 1992, Ap. J. Lett., 396, 6.

3. Bennett, C.L. *et al.* 1996, Ap. J. Lett., 464, 1.
4. Berger, J.O. 1985, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag).
5. Blandford, R.D. & Narayan, R. 1992, Ann. Rev. Astron. Astrophys., 30, 311.
6. Bond, J.R. 1994, in *Proceedings of the IUCAA Dedication Ceremonies*, (ed. T. Padmanabhan; John Wiley & Sons).
7. Bond, J.R. 1995, Phys. Rev. Lett., 74, 4369.
8. Bond, J.R. 1996, in *Cosmology and Large-Scale Structure* (1994 Les Houches summer school; ed., R. Schaefer; Elsevier, in press.)
9. Bond, J.R. & Efstathiou, G. 1984, Ap. J. Lett., 285, 45.
10. Bond, J.R. & Efstathiou, G. 1987, M.N.R.A.S, 226, 655.
11. Bunn, E.F. 1995, *Statistical Analysis of Cosmic Microwave Background Anisotropy*, Ph.D. thesis, Physics Department, U.C. Berkeley (ftp://pac2.berkeley.edu/pub/bunn/thesis).
12. Bunn, E.F., Hoffman, Y., & Silk, J. 1996, Ap. J., 464, 1.
13. Bunn, E.F., Liddle, A., & White, M. 1996, preprint (astro-ph/9607038).
14. Bunn, E.F., Scott, D., & White, M. 1995, Ap. J. Lett., 441, 9.
15. Bunn, E.F. & Sugiyama, N. 1995, Ap. J., 446, 49.
16. Bunn, E.F. & White, M. 1996, preprint (astro-ph/9607060).
17. Bunn, E.F., White, M., Srednicki, M., & Scott, D. 1994, Ap. J., 429, 1.
18. Dodelson, S. & Jubas, J. 1995, Ap.J., 439, 503.
19. Doroshkevich, A.G., Zel'dovich, Ya.B., & Sunyaev, R. 1978, Sov. Astron., 22, 523.
20. Górski, K. 1994, Ap. J. Lett., 430, 85.
21. Górski, K. *et al.* 1994, Ap. J. Lett., 430, 89.
22. Górski, K. *et al.* 1996, Ap. J. Lett., 464, 11.
23. Gull, S.F. & Daniell, G.J. 1978, Nature, 272, 686.
24. Gunn, J.M. & Peterson, B.A. 1965, Ap. J., 142, 1663.
25. Hinshaw, C. *et al.* 1996, Ap. J. Lett., 464, 17.
26. Hu. W. 1995, *Wandering in the Background: A Cosmic Background Explorer*, Ph.D. Thesis, Physics Department, U.C. Berkeley (ftp://pac2.berkeley.edu/pub/hu/thesis).
27. Hu, W. 1996, in *The Universe at High-z, Large-Scale Structure and the Cosmic Microwave Background*, eds. E. Martinez-Gonzalez & J.-L. Sanz (Springer-Verlag) (astro-ph/9511130).
28. Hu, W., Scott, D., & Silk, J. 1994, Phys. Rev. D, 49, 648.
29. Hu, W., Scott, D., Sugiyama, N. & White, M. 1995, Phys. Rev. D., 52, 5498.
30. Hu, W. & Sugiyama, N. 1994, in *CWR CMB Workshop: Two Years After COBE*, eds. L. Krauss, P. Kernan (World Scientific, Singapore, p. 188).
31. Hu. W. & Sugiyama, N. 1995a, Phys. Rev. D, 51, 2599.
32. Hu, W. & Sugiyama, N. 1995b, Ap. J., 444, 489.
33. Hu. W. & Sugiyama, N. 1996, preprint (astro-ph/9510117).
34. Hu, W., Sugiyama, N., & Silk, J. 1996, Nature, in press (astro-ph/9604166; see also http://www.sns.ias.edu/~whu/physics/physics.html).
35. Hu, W. & White, M. 1996, in *Proceedings of the XXXIst Moriond Meeting, Microwave Background Anisotropies*, in press.
36. Jackson, J.D. 1975, *Classical Electrodynamics* (Wiley).
37. Jungman, G., Kamionkowski, M., Kosowsky, A., & Spergel, D. 1995, preprint, astro-ph/9512139.
38. Karhunen, K. 1947, *Über Lineare Methoden in der Wahrscheinlichkeitsrechnung* (Kirjapaino oy. Sana, Helsinki).
39. Kosowsky, A., Kamionkowski, M., Jungman, G., & Spergel, D.N. 1996 preprint (astro-ph/9605147).
40. Lineweaver, C.H. *et al.* 1994, Ap. J., 436, 452.
41. Netterfield, C.B., Devlin, M.J., Jarosik, N., Page, L., & Wollack, E.J. 1996, preprint

(astro-ph/9601197).
42. Ostriker, J. & Vishniac, E. 1986, Ap. J., 306, 51.
43. Peebles, P.J.E. 1987, Nature, 327, 210.
44. Peebles, P.J.E. & Yu, J.T. 1970, Ap. J., 162, 815.
45. Press, W.H. 1996, in *Unsolved Problems in Astrophysics* (ed. J.P. Ostriker, Princeton University Press, in press, astro-ph/9604126).
46. Rees, M. & Sciama, D. 1968, Nature, 519, 611.
47. Rice, J.A. 1995, *Mathematical Statistics and Data Analysis* (Duxbury).
48. Rybicki, G. & Press, W. 1992, Ap. J. Lett., 432, 75.
49. Sachs, R.K. & Wolfe, A.M. 1967, Ap. J., 147, 73.
50. Scaramella, R. & Vittorio, N. 1993, M.N.R.A.S, 263, 17.
51. Scott, D., Silk, J. & White, M. 1995, Science, 268, 829.
52. Seljak, U. 1994, Ap. J. Lett., 435, 87.
53. Seljak, U. 1996a, Ap. J., 460, 549.
54. Seljak, U. 1996b, Ap. J., 463, 1..
55. Seljak, U. & Bertschinger, E. 1994, Ap. J. Lett., 417, 9.
56. Seljak, U. & Zaldarriaga, M. 1996, preprint (astro-ph/9603033).
57. Silk, J. 1967, Nature, 215, 1155.
58. Silk, J. 1968, Ap. J. Lett., 151, 459.
59. Silk, J. 1982, Acta Cosmologica, 11, 75.
60. Smoot, G. *et al.* 1992, Ap. J. Lett., 396, 1.
61. Sunyaev, R.A. 1977, Sov. Astron. Lett., 3, 491.
62. Sunyaev, R.A. & Zel'dovich, Ya.B. 1970, Astrophys. Space Sci., 7, 3.
63. Sunyaev, R.A. & Zel'dovich, Ya.B. 1972, Comm. Astrophys. Space Phys., 4, 73.
64. Tegmark, M. 1995, Ap. J., 455, 429.
65. Tegmark, M. 1996a, M.N.R.A.S, 280, 299.
66. Tegmark, M. 1996b, Ap. J. Lett., 464, 35.
67. Tegmark, M. 1996c, to appear in Proc. Enrico Fermi, Course CXXXII, Varenna (astro-ph/9511148).
68. Tegmark, M. & Bunn, E.F. 1995, Ap. J., 451, 1.
69. Tegmark, M. & Efstathiou, G. 1996, preprint (astro-ph/9507009).
70. Tegmark, M., Taylor, A., & Heavens, A. 1996, preprint (astro-ph/9603021).
71. Viana, P.T.P. & Liddle, A. 1996, M.N.R.A.S, in press (astro-ph/9511007).
72. Vittorio, N. & Silk, J. 1984, Ap. J. Lett., 285, 39.
73. Vogeley, M.S. & Szalay, A.S. 1996, Ap. J., 465, 34.
74. White, M. & Bunn, E.F. 1995, Ap. J., 450, 477.
75. White, M., Scott, D., & Silk, J. 1994, Ann. Rev. Astron. Astrophys., 32, 319.
76. White, M. & Srednicki, M. 1995, Ap. J., 443, 6.
77. Wiener, N. 1949, *Extrapolation and Smoothing of Stationary Time Series* (Wiley).
78. Wilson, M. 1983, Ap. J., 273, 2.
79. Wilson, M. & Silk, J. 1981, Ap. J., 243, 14.
80. Wright, E.L. *et al.* 1992, Ap. J. Lett., 396, 11.
81. Wright, E.L. *et al.* 1994a, Ap. J., 420, 1.
82. Wright, E.L. *et al.* 1994b, Ap. J., 436, 443.
83. Zel'dovich, Ya.B. & Sunyaev, R., Astrophys. Space Sci., 4, 301 (1969).